

June 16, 2017  
DRAFT

# **Characterizing Guitar Strings as Inharmonicity Trajectories for Automatic Tablature Transcription**

Jonathan Michelson

May 2017

School of Music  
Carnegie Mellon University  
Pittsburgh, PA 15213

**Thesis Committee:**

Dr. Richard Stern, Dr. Thomas Sullivan

*Submitted in partial fulfillment of the requirements  
for the degree of Master of Science in Music and Technology*

June 16, 2017  
DRAFT

## **Abstract**

We propose and evaluate an additional inharmonicity-based method for automatic tablature transcription using only audio. First, we characterize each guitar string as the linear relation between its log-inharmonicity and fundamental pitch. Then we use these learned trajectories or regressions to classify string labels of unseen guitar notes: the system assigns to a pitch the label of the line which approximates the note’s estimated inharmonicity with minimal residual. Given the tuning of the guitar that produced the unseen note, we infer its associated tablature, which allows for further string classification refinement by rejecting and reassigning implausible fret assignments. Average F1-scores achieved with this system for classical, acoustic, and electric guitars in the RWC instruments dataset were 0.92, 0.94, and 0.79 respectively, rivaling performance of existing inharmonicity-based systems. We also derive and evaluate an inharmonicity compensation feature to allow for transcription of arbitrarily-tuned guitars, provided the tuning is known. With this enhancement, our average system accuracy improves by 7% on alternate-tuned guitars, compared to performance with no tuning compensation.

June 16, 2017  
DRAFT

June 16, 2017  
DRAFT

### **Acknowledgments**

Rich Stern Tom Sullivan Che-yuan Liang, Raymond Xia Hodgkinson – for sharing MAT algorithm Abesser,Barbancho – for their helpful correspondence

June 16, 2017  
DRAFT

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background</b>	<b>5</b>
2.1	Previous Tablature Transcription Work . . . . .	5
2.1.1	Graph Search Methods . . . . .	5
2.1.2	Machine Learning Solutions . . . . .	5
2.1.3	Multimedia Approaches . . . . .	6
2.1.4	The Partial Coincidence Tally (PCT) Method . . . . .	6
2.2	Inharmonicity Motivation . . . . .	7
2.3	Previous Inharmonicity Estimation Work . . . . .	8
<b>3</b>	<b>Characterizing Strings as Inharmonicity Trajectories</b>	<b>11</b>
3.1	Motivation . . . . .	11
3.2	Inharmonicity Estimation . . . . .	14
3.3	Log-inharmonicity . . . . .	15
3.4	Tuning Compensation . . . . .	19
3.5	Learning Inharmonicity Trajectories . . . . .	22
3.6	String Classification . . . . .	24
3.7	Tablature Conversion and Refinement . . . . .	27
3.8	System Overview . . . . .	28
<b>4</b>	<b>Experiments and Results</b>	<b>29</b>
4.1	Benchmark RWC Evaluation . . . . .	29
4.2	Tuning Compensation . . . . .	31
<b>5</b>	<b>Discussion</b>	<b>35</b>
5.1	Analysis . . . . .	35
5.2	Threats to Validity . . . . .	43
5.3	Future Work . . . . .	44
<b>6</b>	<b>Conclusion</b>	<b>47</b>
	<b>Bibliography</b>	<b>49</b>

June 16, 2017  
DRAFT



# List of Figures

1.1	Musical score and two identical realizations of that score in tablature notation. Figure taken from [6]. . . . .	2
2.1	Power spectrum of note D2 plucked on an electric guitar at string 5 and fret 5. Red dashed lines are drawn at integer multiples of the fundamental. Note the rightward skew of the measured partial peaks. . . . .	8
3.1	Example inharmonicity trajectories of acoustic guitar strings. Only frets 0 (open-string) through 3 are shown. Observe the restriction to the exponential trajectory defined by equation (3.1). . . . .	12
3.2	Median-adjustive-trajectory (MAT) method flowchart. . . . .	16
3.3	Inharmonicity estimates for frets 0-12 on strings 1-6 of RWC acoustic guitar 111AG. Note the exponential trajectory as in Fig. 3.1 and the increased segregation since notes and their inharmonicities are now plotted against their fundamental instead of their fret number. . . . .	17
3.4	Log-inharmonicity estimates for same frets, strings, and guitar as Fig. 3.3 . . . .	18
3.5	Effect of alternate tuning on an electric guitar's log-inharmonicities. The circles are log-inharmonicity estimates of frets 0-12 on string 6 of RWC electric guitar EG131 in standard tuning, and the dashed line is their bisquare-weighted regression. The asterisks are log-inharmonicity estimates for the same frets, string, and guitar, but tuned two semitones down. The dotted line passing through them is the <i>predicted</i> regression obtained by applying equation (3.20) to the dashed line. .	21
3.6	Comparison of string characterization approaches for our recorded Fender Telecaster. Top: inharmonicity <i>trajectories</i> . Note how each line begins precisely at the first marker of each string (i.e. the open-string note) and continues along a trajectory that ignores empirical inharmonicity estimates. String 1 most clearly shows this. Additionally, the trajectories for strings 3 and 4 are effectively colinear and therefore indiscriminible. Bottom: inharmonicity <i>regressions</i> . Each regression was obtained by minimizing the bisquare weighted error of the string's residuals, so the lines more closely follow the empirical inharmonicity scatterplot estimates. . . . .	23

3.7	Classifying two new notes A and B using learned log-inharmonicity characterizations. Dashed lines are actual learned regressions for RWC acoustic guitar AG091. Note A with fundamental pitch 62 is assigned to string 3, since its line characterization is closest to A's inharmonicity estimate. The residual of note B, with fundamental pitch 76, is smallest with respect to string 1, so it's assigned to that string under the least-square-error decision of equation (3.24). . . . .	26
5.1	Superimposed string-wise regressions of all three RWC classical guitars. Top: strings 1, 2, and 3. Bottom: strings 4, 5, and 6. Separated for visibility. Axes scales identical for comparison. Observe the consistent placement and orientation despite varying guitar. . . . .	37
5.2	. . . . .	38
5.3	. . . . .	39
5.4	. . . . .	41
5.5	Boxplots of log-inharmonicity regression residuals for EG131 by dynamic level. Group 1: piano, Group 2: mezzo, Group 3: forte. Red center horizontal lines are medians; top and bottom edges of blue boxes represent upper and lower quartiles; whiskers depict full range of data excluding outliers; red plus signs denote outliers. Vertical axes not constant throughout plots; we're highlighting relative distributions of the boxplots here. . . . .	42
5.6	Solid black lines: probability of correctly transcribing a note given the residual width under which the note's log-inharmonicity falls, measured by the left-hand-side vertical axes. Dashed orange lines: CDF of the log-inharmonicity residuals, measured by the right-hand-side vertical axis. . . . .	46

## List of Tables

4.1	Error probability comparisons between PCT, residual minimization with theoretical trajectories, and residual minimization with empirical trajectories. PCT results are taken from [6]; we've excluded error due to $f_0$ estimation in their results.	31
4.2	F1-scores for theoretical and empirical log-inharmonicity trajectory classification. PCT column removed because F1-scores weren't reported in [6]. . . . .	32
4.3	String-wise F1-scores for the RWC classical guitars. . . . .	32
4.4	String-wise F1-scores for the RWC acoustic guitars. . . . .	32
4.5	String-wise F1-scores for the RWC electric guitars. . . . .	33
4.6	Best string-wise F1-scores for our Fender Telecaster at various tunings. Regressions trained on all three RWC electric guitars. WSU: whole-step up (from standard); WSD: whole-step down (from standard); orig.: no tuning compensation used in this trial; comp.: tuning compensation used in this trial. The "Overall" row displays average F1-score of each column. . . . .	33
5.1	Confusion matrix for one classification trial with EG131. Strings 3 and 4, as they are here, were frequently misclassified because of their regressions' poor discriminability. . . . .	40

June 16, 2017  
DRAFT

## Chapter 1: Introduction

The guitar is a popular musical instrument whose family comprises a diverse collection of stringed instruments. Its most prominent classes are the acoustic, classical, and electric guitars, and intraclass variation in characteristics such as body shape, string material, and amplification mode is high. Despite this diversity, the majority of guitar configurations have six strings tuned to *E*2 (82Hz), *A*2 (110Hz), *D*3 (147Hz), *G*3 (196Hz), *B*3 (247Hz), and *E*4 (330Hz). To play, a performer presses these strings with one hand against the neck’s metal frets, each of which increase the string’s pitch by one half-step, while strumming or plucking them with the other hand. Guitars typically have upwards of 20 frets, allowing for versatile musical passage realizations along the fretboard.

A consequence of this is liberal pitch overlap between neighboring strings. Conventional music scores that represent passages as notes and chords therefore fail to communicate fretboard position. This ambiguity renders scores unfit for guitar students trying to learn more skillful placement of scales and arpeggios, or for enthusiasts trying to decrypt the fretboard positions used on recordings of virtuosic players’ riffs. Figure 1 illustrates this ambiguity.

Tablature, on the other hand, is an alternative music notation that doesn’t suffer from the one-to-many mapping of scores. The staff, instead of representing pitch as in classical scores, depicts a birds-eye view of the guitar neck from the performer’s perspective. Each of the six horizontal lines signify a corresponding string on the guitar, and numbers on each line specify the fret to be played. Time progresses from left to right as in scores.

Tablature (tabs for short) is widely popular for a number of likely reasons. Its accessibility



Figure 1.1: Musical score and two identical realizations of that score in tablature notation. Figure taken from [6].

and intuition are attractive to those without formal musical training; learning from a pictorial representation of one’s instrument must seem more attractive to beginners than learning musical theory foundations and decoding scores. And because tabs are easily encodable and interpretable with simple ASCII characters, authoring them isn’t limited to those with knowledge of any sort of tab-specific generation program, unlike with scores. [17].

Accurate tablature transcription is a fairly tedious task, however. Dependable tabs containing few errors typically require a seasoned musician’s careful listening to an audio recording, and possibly video recordings if available too. Reliable automation of this transcription task would benefit the guitar student community by expediting quality tab generation and unearthing fretboard positions of guitar riffs in songs for which accurate tabs don’t yet exist. Recent technologies proposed to encourage proliferation of accessible tab transcription systems, such as web framework Robotaba [8], demonstrate a desire among MIR researchers to see this theoretically mature but practically nascent field mature into student-benefitting applications. Closely related to tab transcription is the task of string classification: producing the correct tablature is straightforward if the guitar’s tuning is known and the correct string labels are identified.

In the next chapter, we survey previous work in the field of automatic tablature transcription, and highlight the feature known as inharmonicity. Then in Chapter ??, we motivate inharmonic-

June 16, 2017

DRAFT

ity's suitability for the string discrimination task, looking at Barbancho's work as an example. We transition to discussing the novel approach of our system. Chapter 4 contains descriptions and results of our experiments, and the discussion Chapter ?? regards an analysis of the results, shortcomings, and remaining work. We summarize this work in our final chapter.

June 16, 2017  
DRAFT



## **Chapter 2: Previous Work**

### **2.1 Automatic Tablature Transcription**

#### **2.1.1 Graph Search Methods**

A popular framework for the stringed-instrument transcription task, introduced by Sayegh [?] for both transcription and fingering problems, is that of weighted directed graphs that capture all plausible sequences of fretboard positions for a given musical passage. Edge weights in the graph are typically informed by mechanical or musical desirability of the fretboard sequences implied by the vertex transitions. Radicioni [?] used this graph paradigm to obtain optimal tablatures from music scores. Yaezawa [?] extracted pitch and melody information from guitar recordings, and computed an optimal fingering based on plausible multipitch estimation results. Burlet [?] developed a polyphonic extension to monophonic optimal search algorithms. Radisavljevic [18] proposed a method to learn the optimal weights on the graph given optimal tablature.

#### **2.1.2 Machine Learning Solutions**

A number of teams have used machine learning techniques to ascertain fretboard position or related parameters. Gagnon [1] proposed using neural networks on scores to deduce general fretboard region position and number of strings playing, while Tuohy [?] used a genetic algorithm on scores to obtain tablature. Barbancho et al. [5] attempted guitar string classification with a Fisher linear discriminant using a multitude of time and frequency-domain features. Abesser [4],

Dittmar [9], and Kehling [16] achieve good string transcription performance after training an SVM on standard-tuning guitar audio and using a slew of partials-related features. Burlet [] developed a deep belief network that achieved fretboard transcription from audio.

### 2.1.3 Multimedia Approaches

Some researchers have explored usage of media other than audio or musical information (i.e., scores). Ogrady [] exploited hexaphonic pickups and matrix factorization of their output to record fretboard positions. Paleari [], Burns [], and Kerdvibulvich [] extract tablature from video recordings using image processing and computer vision techniques.

### 2.1.4 The Partial Coincidence Tally (PCT) Method

One system in particular served as inspiration for this work. Barbancho [6] realized the discriminative potential of a singular feature known as inharmonicity, and devised an algorithm that exclusively used inharmonicity to achieve impressive transcription results using only audio. We discuss inharmonicity in detail and methods for its estimation in the next two sections. Barbancho’s method, which we refer to as the ”partial coincidence tally” method, is also given a more thorough treatment later in Chapter ??.

We were particularly interested in the PCT method for a number of reasons. Firstly, its input is audio. This is important because other media are less —; scores are inaccessible to novices, while video recordings somewhat nullify the need for tablature ascertainment in the first place. Among the most interesting use cases for an automatic tab transcription system are certainly direct audio-to-tab capability. Other inputs, like scores or video footage of the guitar neck, are already levels of transcription that detracts from the novelty of such a system. Secondly, PCT is rooted in physics which allows for potential transcription of theoretically any fretboard position. Lastly, their method exclusively uses inharmonicity. Other systems that use inharmonicity [4, 5, 9, 16] exploit additional features that are, in light of [6], superfluous regarding the discriminative

information they contribute.

## 2.2 Inharmonicity Motivation

To understand the motivation for this work, we first need to address the motivation behind inharmonicity as a feature itself. Inharmonicity is key in many string discrimination systems [4, 6, 9, 16]. When an ideal string fixed at both ends is displaced, the restoring force that causes it to oscillate is its tension. However, for a real string with stiffness, its elasticity contributes to this restoring force [11] which causes the frequencies of the modes of vibration to no longer be integer multiples of the fundamental. Instead, they're skewed upward according to:

$$f_k = kf_0\sqrt{1 + \beta k^2} \quad (2.1)$$

where  $f_k$  is the  $k$ th harmonic of fundamental  $f_0$  and  $\beta$  is the string's inharmonicity, defined by

$$\beta = \frac{\pi^3 Q d^4}{64 T l^2}. \quad (2.2)$$

In words, the inharmonicity  $\beta$  of a vibrating string is a dimensionless quantity that depends on the string's Young's modulus  $Q$ , diameter  $d$ , tension  $T$ , and vibrating length  $l$ , and which scales the degree of deviation of the string's  $k$ th partial according to equation (2.1). See Figure 2.2. Note that for an ideally harmonic string,  $\beta = 0$  and (2.1) reduces to  $f_k = kf_0$ , aligning with intuition about harmonics' ideal locations in frequency as simply integer multiples of the fundamental.

Inharmonicity's suitability for the string discrimination task can be understood through equation (2.2). The six strings of any guitar necessarily exhibit different combinations of  $Q$ ,  $d$ , and  $T$  due to varying thicknesses, material, and tuning. Each string therefore has an "inharmonic signature" which presumably distinguishes it from the others. Capturing the pattern of a string's inharmonicity variation along its frets should thus reveal sufficiently identifying information.

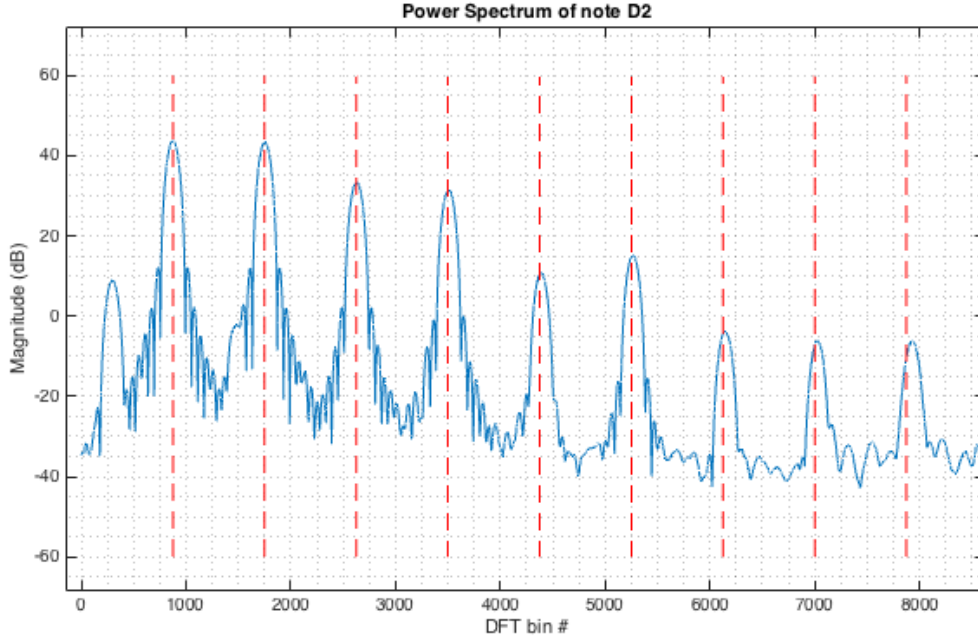


Figure 2.1: Power spectrum of note D2 plucked on an electric guitar at string 5 and fret 5. Red dashed lines are drawn at integer multiples of the fundamental. Note the rightward skew of the measured partial peaks.

## 2.3 Inharmonicity Estimation

Empirical estimation of inharmonicity is itself a topic of research which merits a brief review. Using pitch extraction techniques to perform estimation was first attempted by Galembo in 1979 and 1986 [12, 13]. Several additional methods have been proposed since then. In 1994, Galembo [2] hypothesized a connection between partials-based fundamental pitch estimates and the degree of inharmonicity in a spectrum. They specifically investigated cepstral analysis and a variant of the harmonic product spectrum method applied to both synthetic tones and recorded piano notes. They found that, because these two techniques leverage periodicity in the frequency domain to produce their pitch estimates, inharmonicity influenced the quality of their estimates in a deterministic manner: more inharmonicity produced wider, less focused pitch estimates lobes, and from quantifying this relation an inharmonicity could be obtained.

The same authors in [3] introduced another method in 1999, dubbed the inharmonic comb filter (ICF) method. To estimate a note’s inharmonicity, they first applied sets of comb filters to the note’s spectrum. The locations of the comb filters’ notches were parametrized with a range of inharmonicity values that sufficiently sampled the note’s expected inharmonicity range. The ICF whose inharmonicity best matched that of the note would produce the output with lowest spectral energy, since its notches best aligned with the inharmonic note partials. The inharmonicity parameter associated with this output-minimizing ICF was thus selected as the estimate.

Rauhala [19] introduced an efficient procedure in 2007 that iteratively catalogued estimates of the partial frequencies’ deviations and returned increasingly better estimates of the inharmonicity. The algorithm, aptly referenced as the partial frequency deviations (PDF) method, is initialized with a reasonable first estimate of inharmonicity, and then the note’s spectrum is searched for partial peaks. Differences are measured between the locations of these discovered peaks and those of the expected peaks, and the aggregate deviation trend is used to refine the inharmonicity estimate – a majority of positive differences implies the inharmonicity should be reduced, and a majority of negative differences implies the opposite.

In 2009, Hodgkinson [15] proposed a yet more efficient and accurate algorithm, named median-adjustive trajectories (MAT). Their routine exploits the fact that inharmonicity can be estimated with any two partials and their corresponding indices in the harmonic series. Pairs of low-index partials are first considered, since they’re minimally affected by inharmonic skew and therefore have high location reliability. With these initial inharmonicity estimates, additional partials are collected and used to refine the previous inharmonicity estimates, with which more partials are located, so on and so forth. MAT was the inharmonicity estimation routine employed in this work, and is discussed at length in Chapter 4.

Three years later, Barbancho [6] introduced another PFD-based approach as part of a guitar tablature transcription system. They started by cataloguing the first 10 partials’ deviations, then fitting a polynomial to the PFD curve. They derived the relation between the polynomial’s co-

efficients and the inharmonicity, and returned an initial inharmonicity estimate according to this derivation. Then using this inharmonicity to aid in their search for higher-index partials, they catalogued a larger number of partials' deviations than that of the first iteration, and performed the same polynomial fit and inharmonicity derivation routine. This would refine the inharmonicity estimate, allowing for increasing numbers of partials for which to search.

Also in 2012, Abesser [4] used parametric spectral modeling, again in the context of a guitar tablature transcription system. To estimate the inharmonicity of a plucked guitar note, they modeled each audio frame with the autoregressive (AR) filter that best approximated its spectral content. They obtained the poles' locations and the ideal harmonics' locations, and like Barnbancho [6] fit the PFD curve with a polynomial from whose coefficients an inharmonicity estimate could be derived.

## Chapter 3: Characterizing Strings as Inharmonicity Trajectories

In this chapter, we present our automatic tablature transcription system in detail. We begin with the motivation for our approach, then present its various building blocks: the specific inharmonicity estimation algorithm we used, transformation of inharmonicity to log-inharmonicity, compensation for arbitrary tuning, learning log-inharmonicity trajectories, string classification using the learned trajectories, and the final tablature conversion and refinement process.

### 3.1 Motivation

Barbancho [6] derived that inharmonicity varies in a deterministic manner with respect to fret positions along a given string. Specifically, they showed that the inharmonicity  $\beta(s, n)$  of a note produced by string  $s$  at fret number  $n$  could be expressed in terms of the inharmonicity of the open-string note  $\beta(s, 0)$  according to:

$$\beta(s, n) = \beta(s, 0)2^{\frac{n}{6}} \quad (3.1)$$

In other words, the variation in inharmonicity along a given string can be expressed simply in terms of its open-note inharmonicity, effectively defining a restricted trajectory along which the inharmonicity must vary. Figure 3.1 illustrates these trajectories.

Recognizing this, they were able to successfully transcribe string and fret number of an un-

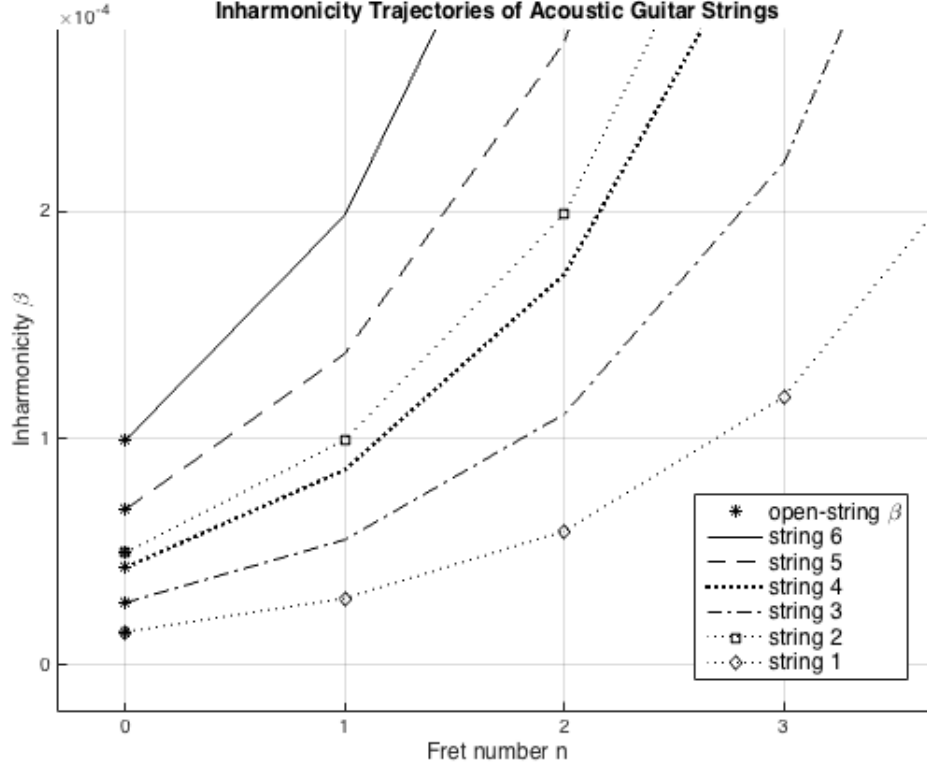


Figure 3.1: Example inharmonicity trajectories of acoustic guitar strings. Only frets 0 (open-string) through 3 are shown. Observe the restriction to the exponential trajectory defined by equation (3.1).

known guitar note. Their system’s essential mechanism was a similarity-determination routine  $A(X, \hat{\beta}_s)$  that quantified the resemblance between an unknown note’s spectrum  $X$  and the expected inharmonicities  $\hat{\beta}_s$  of candidate strings  $s$  from which the note could have originated. First, to estimate an open-note’s inharmonicity they followed the procedure described in section 2.3 in which they estimated its fundamental, then catalogued its partials’ deviations from their ideal harmonic locations, and fit to them a polynomial from whose coefficients they derived the inharmonicity estimate. Next, they inferred the corresponding fret on which the note could be played for each candidate string, and used these string-fret candidate numbers in equation (3.1) to obtain expected inharmonicity values of the unknown note for each candidate string. In other words, they followed each candidate string’s inharmonicity trajectory to arrive at an expected



inharmonic value at the appropriate fret. Lastly, they applied their similarity determination algorithm  $A(X, \hat{\beta}_s)$ , which went as follows. For each candidate fretboard position, they obtained the expected skewed frequency locations at which to find the inharmonic peaks, and searched the note’s spectrum for them using search windows centered at those expected locations. They performed this search for a set number of peaks, all the while cataloguing their deviation from the search window centers. Their motivation was that the expected inharmonicity of the true fretboard position would produce a spectrum with inharmonic peaks that were measurably more coincident with the inharmonic peaks of the unknown note than were any of the imposter strings. They found that, indeed, the inharmonicity of the fretboard position on which the note was truly plucked yielded the largest number of coincident located peaks inside the spectral search window bounds, hence our name “partials coincidence tally” for their algorithm. Their work demonstrated the sufficiency of inharmonicity as a discriminating feature.

This work capitalizes further on their trajectory realization. Rather than zooming into fretboard positions’ spectra to evaluate their candidacy, we propose characterizing each candidate string by its inharmonicity trend and classifying a note as belonging to the string whose inharmonicity trajectory evaluated at the note’s fundamental best approximates the note’s inharmonicity. In other words, our similarity routine is simply some error calculation  $A'(\beta, \hat{\beta})$ , discussed in detail in the next sections, between the note’s inharmonicity  $\beta$  and each candidate fretboard positions’ predicted inharmonicity  $\hat{\beta}$  for the note. This is an algorithmically simple approach, requiring merely inharmonicity estimation of the unknown note, inharmonicity trends of training guitars’ strings, and an error measure. Additionally, this method is governed by the intuitive spatial constraint of inharmonicity trajectory: notes played on a given string can exhibit only a set of inharmonicities, so classifying based on a spatial metric like minimal residual from a regression makes physical sense.

## 3.2 Inharmonicity Estimation

The first step in our system is estimation of the inharmonicity of an unseen note. The algorithm we employed was the median adjustive trajectories (MAT) method [15] proposed by Hodgkinson in 2009, and we present it here. At the core of this method is a recasting of the definition of inharmonicity in terms of any two partials and their indices. We can rearrange equation (2.1) to express the fundamental  $f_0$  in terms of the frequency  $f_m$  of the  $m$ th partial and its index  $m$  as

$$f_0 = \frac{f_m}{m\sqrt{1 + \beta m^2}}. \quad (3.2)$$

Next, we can substitute (3.2) back into  $f_0$  in equation (2.1) to obtain

$$f_k = \frac{k f_m}{m\sqrt{1 + \beta m^2}} \sqrt{1 + \beta k^2}. \quad (3.3)$$

Solving for  $\beta$  yields

$$\beta = \frac{(f_k \frac{m}{k})^2 - f_m^2}{k^2 f_m^2 - m^2 (f_k \frac{m}{k})^2}, \quad (3.4)$$

which defines the inharmonicity of a note in terms of an arbitrary two partials  $k$  and  $m$  and their corresponding frequencies  $f_k$  and  $f_m$ .

The MAT algorithm begins by locating partials 1 and 2 based on a user-input estimate of the fundamental. Searching for these partials is straightforward, since the skew effect of inharmonicity is low for small indices. From these locations, (3.4) is used for every combination of the located partials (which for this first iteration, is only one) to obtain an array of  $\beta$  estimates. The median of the  $\beta$  array is returned as the composite  $\beta$  estimate of this iteration, and this median together with the partials' locations are used in (3.2) to obtain an array of revised  $f_0$  estimates that contains as many elements as partials under consideration. Finally, the median of this  $f_0$  array and the median of the  $\beta$  array are used in equation (2.1) to predict the frequency of the next partial for which to search, which is the third in this case. At succeeding iterations, the

lengths of these  $f_0$  and  $\beta$  arrays grow longer (since both the number of partials, and therefore the number of combinations of partials pairs, increases), and the estimates which they contain grow increasingly more accurate. The algorithm is summarized in the flowchart in Figure 3.2, taken from [15]. MAT determines the ideal number of partials for which to search by finding the highest partial index that lies above the average level of the magnitude spectrum, and terminates once this partial index is reached. The result is an efficient and accurate inharmonicity estimation routine.

### 3.3 Log-inharmonicity

Next in our pipeline, we apply a log transformation to the estimated inharmonicity to make its variation with respect to MIDI pitch linear. This has the advantage of using simpler regression models for classification at later stages. If we reconsider equation (3.1) in terms of MIDI pitch number  $m$  instead of fret number  $n$ , we obtain

$$\beta(s, m) = \beta(s, m_{os}) 2^{\frac{m-m_{os}}{6}}, \quad (3.5)$$

where  $m_{os}$  is the open-string pitch of string  $s$ . The exponential trajectory is maintained since incrementing both fret number and MIDI pitch number accomplishes an equivalent increase in fundamental frequency. Figure 3.3 illustrates the trajectories in terms of MIDI pitches.

Now if we consider the log-inharmonicity  $\beta_l(m)$ , we see that

$$\beta_l(m) = \log_2 \beta(s, m) = \log_2 [\beta(m_{os}) 2^{\frac{m-m_{os}}{6}}] \quad (3.6)$$

$$\beta_l(m) = \log_2 \beta(m_{os}) + \frac{m - m_{os}}{6} \quad (3.7)$$

$$\beta_l(m) = (\log_2 \beta(m_{os}) - \frac{m_{os}}{6}) + (\frac{1}{6})m, \quad (3.8)$$

where string  $s$  has been dropped from the notation since we're looking only at variation along a

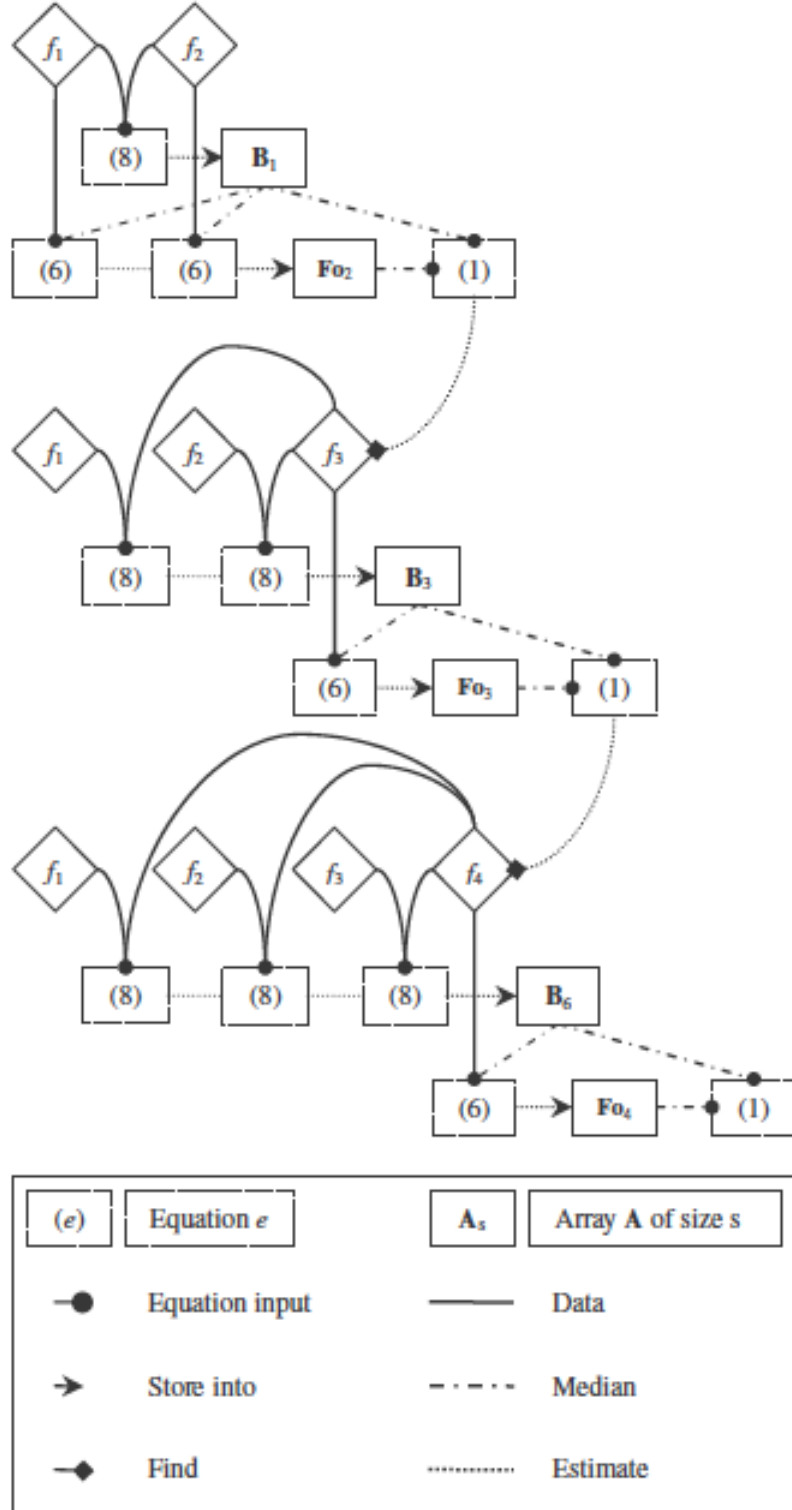


Figure 3.2: Median-adjustive-trajectory (MAT) method flowchart.

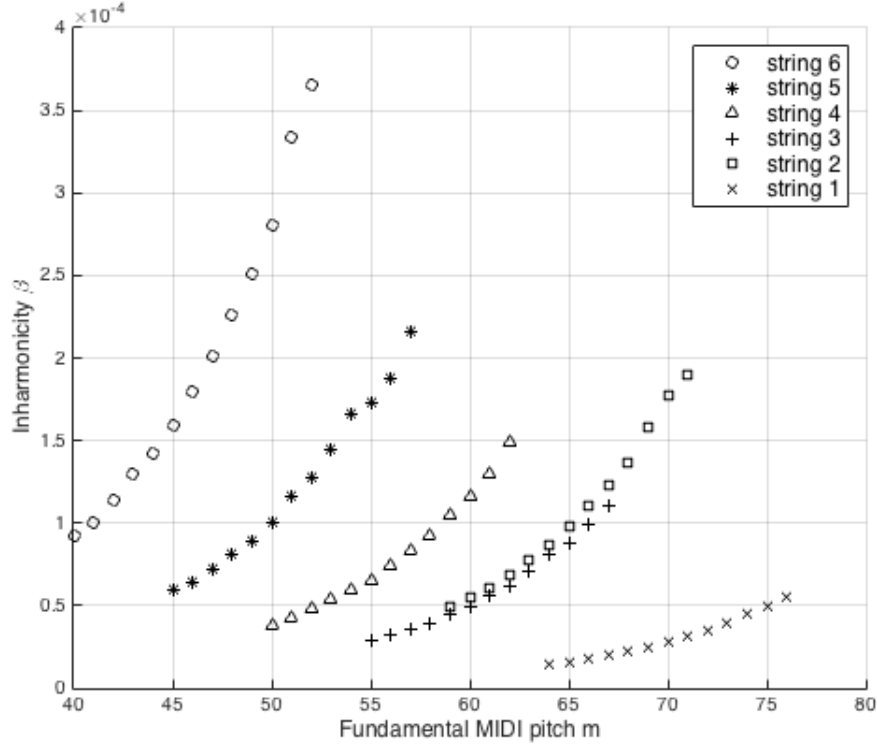


Figure 3.3: Inharmonicity estimates for frets 0-12 on strings 1-6 of RWC acoustic guitar 111AG. Note the exponential trajectory as in Fig. 3.1 and the increased segregation since notes and their inharmonicities are now plotted against their fundamental instead of their fret number.

fixed string. Substituting  $w_0$  for  $\log_2 \beta(m_{os}) - \frac{m_{os}}{6}$  and letting  $w_1 = \frac{1}{6}$ , we obtain the familiar linear form

$$\beta_l(m) = w_0 + w_1 m, \quad (3.9)$$

showing us that the log-inharmonicity trajectory of a given string varies linearly with respect to MIDI pitch. See Figure 3.3. This relationship also holds for fret number, but fundamental pitch is a more likely and generic feature to have been estimated, and consequently is more relevant to consider here.

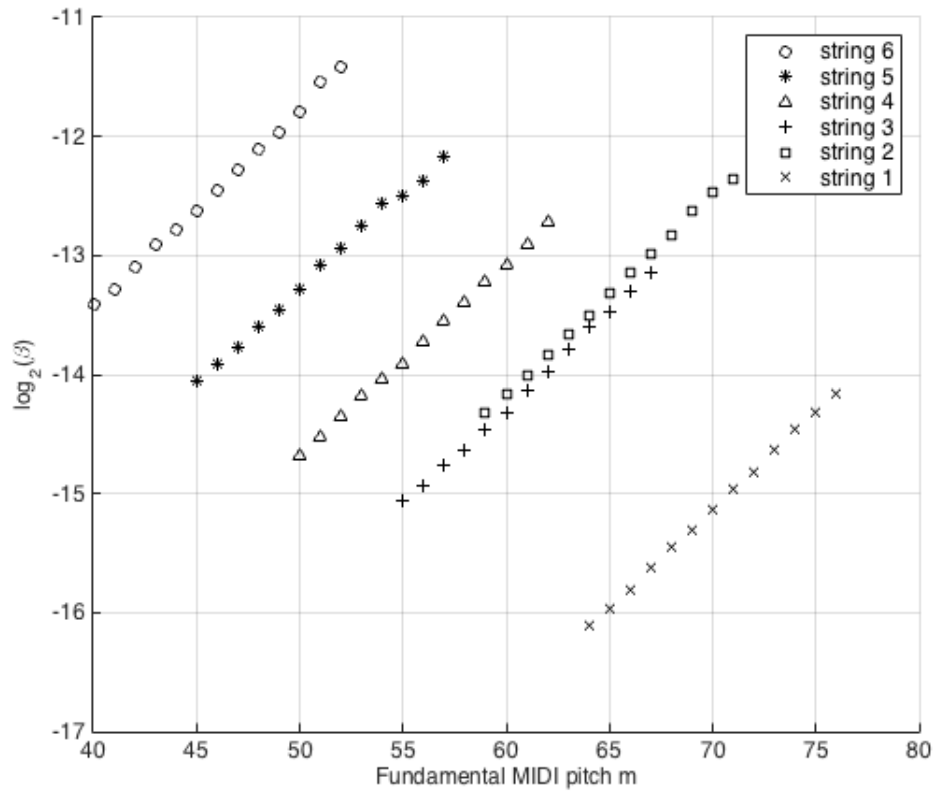


Figure 3.4: Log-inharmonicity estimates for same frets, strings, and guitar as Fig. 3.3

### 3.4 Tuning Compensation

After transforming our data into log-inharmonicity and MIDI pitch space, we apply a tuning compensation feature step if the test guitar is known to be in alternate-tuning. In this section, we derive the straightforward adjustment and explain how it is incorporated into our system.

Though standard tuning is the most common pitch configuration of six-string guitars, there exist numerous other tunings in which performers often play. Aside from altering the musically of the instrument, alternate tunings complicate the transcription process by introducing uncertainty about the open-string pitches. They distort inharmonicity-based methods, since the tensions  $T$  of alternately-tuned guitar strings differ from those in standard-tuned strings, thereby affecting estimation of the inharmonicity.

Current string classification systems don't address this degree of freedom. The method proposed by Barbancho [6] is technically tuning-invariant, but would require access to recordings of the alternately-tuned test guitar's open strings and so isn't considered a classifier here. Other supervised learning systems [4, 9, 16] are trained and evaluated only on standard-tuning guitars.

A possible approach to augment inharmonicity-based string classifiers with tuning-invariant performance is to simply introduce a scaling factor on the expected inharmonicity, or equivalently an additive factor on the expected log-inharmonicity. The fundamental frequency  $f_0$  of an ideal vibrating string is related to its tension  $T$  according to

$$f_0 = \frac{1}{2L} \sqrt{\frac{T}{\mu}} \quad (3.10)$$

where  $L$  is string length and  $\mu$  is its density. From this, we can see that

$$T \propto f_0^2. \quad (3.11)$$

Recognizing that for a change in fundamental pitch of  $\Delta m$  semitones, the equivalent change in

frequency is  $2^{\frac{\Delta m}{12}}$ , we see that the proportional change in tension (with all other factors constant) is

$$T \propto 2^{\frac{\Delta m}{6}} f_0^2. \quad (3.12)$$

The resulting log-inharmonicity of this new open-string note with pitch  $m'_{os} = m_{os} + \Delta m$ , with original open-string pitch  $m_{os}$ , is therefore

$$\log_2 \beta(m'_{os}) = \log_2 \left[ \frac{\pi^3 Q d^4}{64 T l^2} (2^{-\frac{\Delta m}{6}}) \right]. \quad (3.13)$$

$$\log_2 \beta(m'_{os}) = \log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{\Delta m}{6}. \quad (3.14)$$

Substituting (3.14) into  $w_0$  from equation (3.9) to obtain the new intercept term  $w'_0$ , we get

$$w'_0 = \log_2 \beta(m'_{os}) - \frac{m'_{os}}{6} \quad (3.15)$$

$$w'_0 = \left( \log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{\Delta m}{6} \right) - \frac{m_{os} + \Delta m}{6} \quad (3.16)$$

$$w'_0 = \log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{m_{os}}{6} - \frac{\Delta m}{3} \quad (3.17)$$

$$w'_0 = w_0 - \frac{\Delta m}{3}. \quad (3.18)$$

The slope coefficient  $w_1 = -\frac{m}{6}$  is independent from  $\Delta m$ , so the affected log-inharmonicity trajectory of this alternately-tuned string is therefore

$$\beta'(m) = w'_0 + w'_1 m \quad (3.19)$$

$$\beta'(m) = w_0 + w_1 m - \frac{\Delta m}{3}. \quad (3.20)$$

In words, equation (3.20) tells us that the effect of a string's alternate tuning is simply addition of a bias term  $-\frac{\Delta m}{3}$  to the log-inharmonicity, where  $\Delta m$  is the semitone deviation from standard



tuning. We should thus be able to compensate our predictive trajectories to arbitrarily-tuned guitars, provided we're given the alternate tuning in which the unseen notes are being performed. As can be seen in Figure 3.4 when one string on the standard-tuned electric guitar is modified using equation (3.20), the predictive alternate-tuning regression aptly captures the trend of the alternate-tuning inharmonicity. This is a straightforward modification for any inharmonicity-exploiting classifier, and we report our results with this modification in the next chapter.

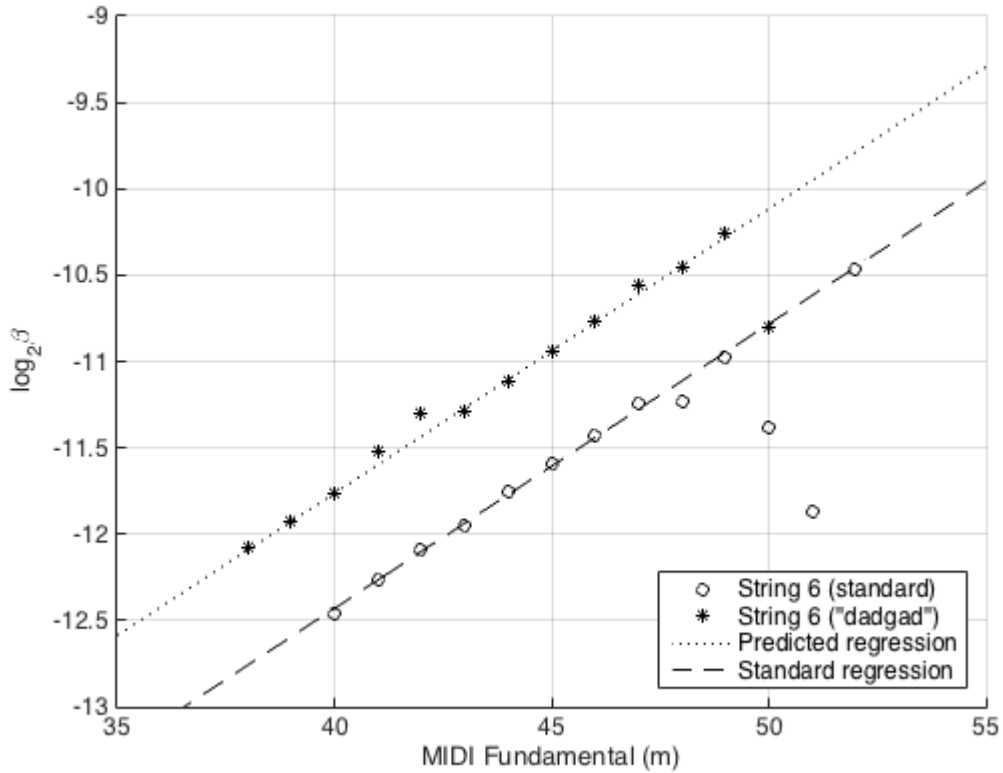


Figure 3.5: Effect of alternate tuning on an electric guitar's log-inharmonicities. The circles are log-inharmonicity estimates of frets 0-12 on string 6 of RWC electric guitar EG131 in standard tuning, and the dashed line is their bisquare-weighted regression. The asterisks are log-inharmonicity estimates for the same frets, string, and guitar, but tuned two semitones down. The dotted line passing through them is the *predicted* regression obtained by applying equation (3.20) to the dashed line.

### 3.5 Learning Inharmonicity Trajectories

We next need to train our system to understand strings' typical log-inharmonicity trajectories. There are two methods we use to do this. The first, which we call *theoretical* trajectory determination, is the method Barbancho employed, so we use this for comparison of our system to theirs. For each guitar, we collect the six notes labeled with fret 0 (i.e. played open-string) from its recordings, and estimate their inharmonicities. To obtain a more robust estimate, we take the string-wise median of these estimates for our final six open-string inharmonicities. Our line trajectories are then simply those defined by equation (3.9).

The other method we use is linear regressions for obtaining *empirical* log-inharmonicity trajectories. This is one of the novel features of this work. Regressing log-inharmonicities of notes against their pitches could capture a more fitting trajectory than the strictly theoretical trajectories obtained from equation (3.9). In this method, we collect all notes with common string labels in our training data, and we perform linear regression of their log-inharmonicities  $\beta$  against their fundamental pitches  $m$  (in MIDI note number format). We do this for each string label. Let  $s \in \{1, 2, 3, 4, 5, 6\}$  be the string label to which each training note is assigned, and  $N_s$  be the number of notes we have belonging to string label  $s$ . If we let  $\mathbf{x}_s^{(i)} = [1, m_s^{(i)}]^T$  represent the  $i$ th note with fundamental pitch  $m^{(i)}$  belonging to string  $s$ , and  $\beta_s^{(i)}$  be the log-inharmonicity of the  $i$ th note belonging to string  $s$ , we can solve

$$\mathbf{w}_s = \arg \min_{\mathbf{w}} \sum_{i=1}^{N_s} (\beta_s^{(i)} - \mathbf{w}^T \mathbf{x}_s^{(i)})^2 \quad (3.21)$$

which yields the weight vector  $\mathbf{w}_s$  that minimizes the sum of squared error between the measured and predicted inharmonicities of the notes belonging to string  $s$ . Figure 3.5 illustrates the differences between the trajectory and regression approaches.

Mistakes in the inharmonicity estimation routine would sometimes produce outliers that deviated substantially from the linear trajectory model, and which disproportionately skewed the re-

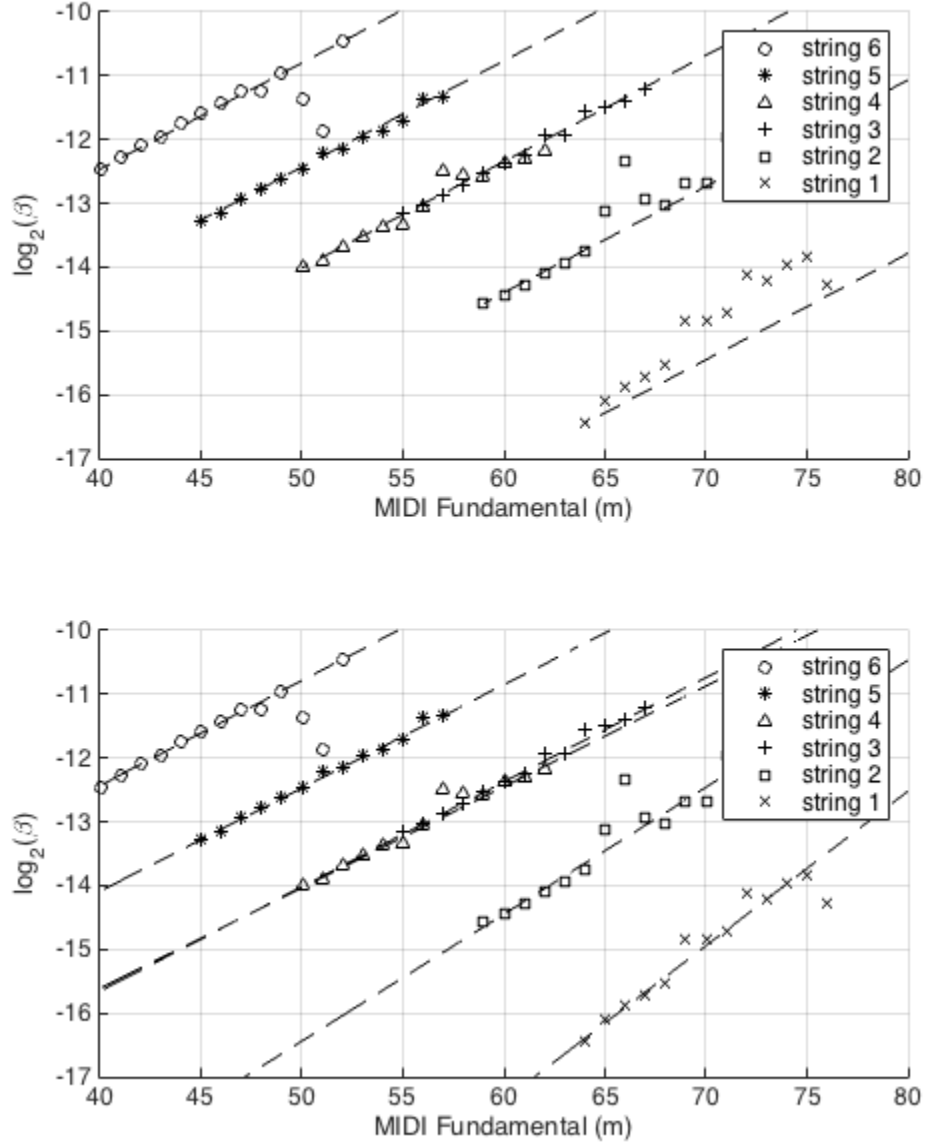


Figure 3.6: Comparison of string characterization approaches for our recorded Fender Telecaster. Top: inharmonicity *trajectories*. Note how each line begins precisely at the first marker of each string (i.e. the open-string note) and continues along a trajectory that ignores empirical inharmonicity estimates. String 1 most clearly shows this. Additionally, the trajectories for strings 3 and 4 are effectively co-linear and therefore indiscriminable. Bottom: inharmonicity *regressions*. Each regression was obtained by minimizing the bisquare weighted error of the string's residuals, so the lines more closely follow the empirical inharmonicity scatterplot estimates.

gressions. For this reason, we rather performed a bisquare weighted linear regression, instead of the standard unweighted regression in (3.21), to empirically approximate the log-inharmonicities. Bisquare weighting is a robust linear regression procedure that minimizes a weighted sum of squares that de-emphasizes outliers. The iterative procedure works as follows. First, a regular least squares fit is obtained. Then, the standardized adjusted residuals are computed according to

$$u = \frac{r_i}{K s \sqrt{1 - h_i}}, \quad (3.22)$$

where  $r_i$  are the plain residuals,  $h_i$  are the leverages,  $K = 4.685$  is a tuning constant, and  $s$  is the robust variance, which is equal to  $M/0.6745$  where  $M$  is the median absolute deviation of the residuals. Next, the robust weights are computed according to

$$w_i = \begin{cases} (1 - u_i^2)^2, & |u_i| < 1 \\ 0, & |u_i| \geq 1 \end{cases} \quad (3.23)$$

The process is repeated, iteratively improving the weights until the fit converges [? ]. The Matlab robustfit routine was used to implement this.

### 3.6 String Classification

The next step is to use these learned string trajectories to classify unknown notes. Since these trajectories tell us how to expect different strings' inharmonicities to vary with respect to pitch, we think of string classification in this context as simply determining which trajectory or line best approximates the unseen note's inharmonicity estimate. To this end, we simply assign a note to the trajectory that minimizes the reconstruction error.

Other decision functions were considered before deciding on residual minimization. We considered using a two-dimensional Euclidean distance metric, but ultimately steered away because it wasn't true to the physical phenomenon – expected inharmonicity on a fixed string

should be a function of only one variable, namely pitch, so incorporating an additional degree of freedom seems inappropriate. Indeed, one can easily imagine situations where the Euclidean distance between a note’s inharmonicity estimate and an erroneous string is smaller than the one-dimensional residual between it and its true string, thereby producing an incorrect string classification.

We also considered a probabilistic decision function. Because, for linear regression, minimizing the residual sum-of-square error of the data is equivalent to maximizing the likelihood of the data if its assumed the data was generated by Gaussian distributions with means centered at the regression line, we also attempted a probabilistic, or ”soft” residual classifier. Unseen notes were assigned to the line from which it was most probably generated, given the lines’ means and variances. The term ”soft” arises from the fact that points are no longer necessarily assigned to the closest line – the decision is evaluated strictly on probabilistic terms. We ultimately avoided this method, however, because our system was suffering from precision issues. The variances of the robust weighted regressions would consistently be tens or hundreds of orders of magnitude smaller than the distances to some outlier inharmonicities. The system would then incorrectly classify these outliers because of insufficient resolution. A workaround wasn’t implemented because this, too, didn’t quite fit the mechanism by which these inharmonicities were being produced in reality. Inharmonicity outliers were strictly due to inharmonicity estimation mistakes, not a physical phenomenon in the guitar, so it didn’t seem appropriate to incorporate this into our model.

We thus settled on minimizing the residual error, which is explained here. We take the  $n$ th note in our test set, characterized by tuple  $(m^{(n)}, \beta^{(n)})$  where  $m^{(n)}$  is its MIDI pitch number and  $\beta^{(n)}$  is its inharmonicity. We’d like to transform this into another tuple  $(s^{(n)}, f^{(n)})$  corresponding to the appropriate string and fret to which this unknown note should be assigned. We straightforwardly assign to  $s^{(n)}$  the index  $j$  of the linear coefficients  $w_j$  whose predicted inharmonicity  $\hat{\beta}_j$

approximates the measured inharmonicity  $\beta^{(n)}$  with least error. More formally, we solve

$$s^{(n)} = \arg \min_j (\beta^{(n)} - \hat{\beta}_j)^2 \quad (3.24)$$

where  $\hat{\beta}_j$  is the inharmonicity prediction of string  $j$ , given by

$$\hat{\beta}_j = \mathbf{w}_j^T \mathbf{x}^{(n)}. \quad (3.25)$$

Here,  $\mathbf{x}^{(n)} = [1, m^{(n)}]^T$  to account for a y-intercept term. Figure 3.6 graphically illustrates this process.

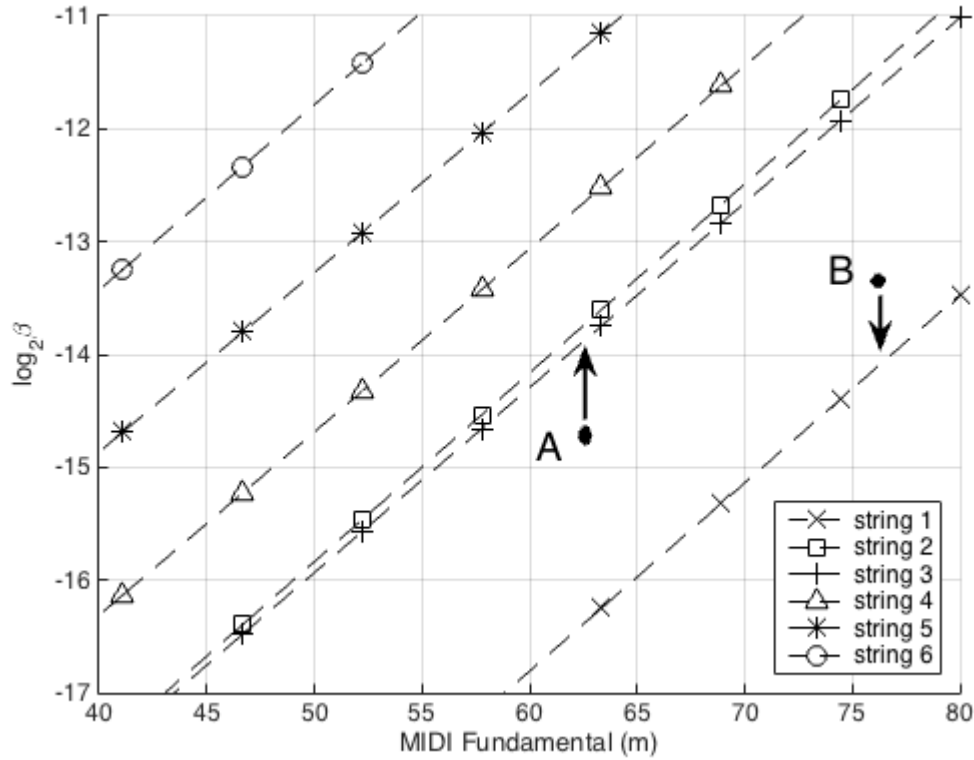


Figure 3.7: Classifying two new notes A and B using learned log-inharmonicity characterizations. Dashed lines are actual learned regressions for RWC acoustic guitar AG091. Note A with fundamental pitch 62 is assigned to string 3, since its line characterization is closest to A’s inharmonicity estimate. The residual of note B, with fundamental pitch 76, is smallest with respect to string 1, so it’s assigned to that string under the least-square-error decision of equation (3.24).

Occasionally, outlier inharmonicity estimates would throw off this classification scheme. To combat this, we employed frame aggregation, a refinement technique used by [4]. In frame aggregation, inharmonicities of multiple audio windows are estimated and classified individually, then the multiple results are aggregated into a compound final decision. In essence, it's a form of averaging that relies on the theoretically low probability that consecutive inharmonicity estimates will all be outliers; the more probable stable estimates will contribute to a more reasonable aggregate string decision. For each frame  $r$  in an  $F$ -frame aggregation of a note, we obtain the string-wise distances  $\mathbf{d}_r = [d_1, d_2, d_3, d_4, d_5, d_6]^T$  between its log-inharmonicity estimate and all possible strings  $s$ . Then, we simply sum the distances  $\mathbf{d}_r$  over all frames and select the string whose aggregate distance is minimal:

$$\hat{s} = \arg \min_{s \in \{1,2,3,4,5,6\}} \sum_{r=1}^F \mathbf{d}_r. \quad (3.26)$$

The result is a tamer inharmonicity estimate and a slight boost in classification accuracy. For our system, we used 5 abutting frames, each 100ms long, which start at the detected onset of a note. Onsets were obtained through automated detection, the details of which are beyond the scope of this work but are cited for reference [7, 10].

### 3.7 Tablature Conversion and Refinement

The final step is transformation of the string classifier output into tablature notation. Because the classifier outputs are string numbers, we still need to infer notes' fret positions. This is a trivial task provided we know the tuning  $\mathbf{t} = [m_1, m_2, m_3, m_4, m_5, m_6]^T$  of the unknown guitar, where  $m_s$  is the MIDI pitch number of the open note on string  $s$ . Because we already have an estimate of the  $n$ th unknown note's MIDI pitch  $m^{(n)}$ , simply taking the difference between  $m^{(n)}$  and its assigned string's open pitch  $m_s$  yields the fret on which the note was played; each fret on a guitar increases the string's pitch by one half-step, or equivalently the MIDI pitch number by

one integer.

We further refine the fretboard estimates with "plausibility filtering" [4] in which we reject and reassign those string decisions that are implausible given the constraints and assumptions of the data. If our system assigns to a note a string on which the corresponding fretted position is negative, a mistake has clearly been made since frets can't possibly be lower than 0, or the open-string. Similarly, if our system assigns to a note a string on which the corresponding fret is greater than 12, we know this is an error as well because our guitar recordings feature only frets 0-12. When either of these situations arise, we reject the string classification decision and select the next closest string. If the succeeding string is also implausible, the process is iterated until all six strings have been considered.

## **3.8 System Overview**



## Chapter 4: Experiments and Results

### 4.1 Benchmark RWC Evaluation

We used subsets of the Real World Corpus Music Instrument Database (RWC-MDB-I) [14] and personal recordings of an electric guitar to conduct our experiments.

Our RWC subset comprised nine guitar recordings (three classical, three acoustic, and three electric), each of which were performed twelve times exhibiting various permutations of particular musicality parameters (playing style, dynamic level, pickup selection). For our work, we used only six of the twelve recordings for each of the electric guitars, as the discarded ones featured musicality parameters like vibrato and palm-muting that weren't featured in the classical and acoustic sets. The performances themselves were simply clean, isolated, monophonic enumeration of every fret (from open-string to 12th fret) on every string (from string 6 to string 1), constituting 78 total note plucks per audio file. The resolution is 16 bits per sample, at 44.1kHz. Labels of strings and frets are thus obtained by their location of occurrence in the recording. The classical guitars recorded were a Stafford, a Sakurai Kohno Professional-J, and a Yuichi Imai YJ-II; the three acoustic guitars captured were an Ovation, a Yamaha APX, and Yairi WY1; the electric guitars featured were a Fender Stratocaster, an Aria PE, and an Ibanez Artcore.

We also recorded a Fender Telecaster at 16 bits and 44.1kHz in the same vein as the RWC database: string and fret enumeration, totaling 78 notes per recording. We captured various tunings: standard, "DADGAD" (in which the open pitches of strings 6 through 1 are given in its name), "WSU" (whole-step up), and "WSD" (whole-step down), and performed each one with a

pick at center pickup orientation and at a similar moderate dynamic level. Hardware included a Focusrite Saffire Pro 24 into a 2011 Macbook Pro running Logic. No other musicality parameters were captured; we focused on tuning variance here.

Our first experiment was a benchmark comparison of our similarity metric (residual error of each string’s trajectory) to Barbancho’s similarity metric (partials coincidence tally, or PCT). This experiment used *theoretical* inharmonicity trajectories. For each guitar in our RWC subset, we obtained log-inharmonicity estimates of its open-string notes from two-thirds of its recordings, then performed string classification on the remaining third using residual error minimization on the theoretical trajectories, as discussed in Chapter 3. We repeated this twice with different folds for three-fold cross validation, and averaged each fold’s probability of an erroneous transcription to produce composite error probabilities. These are reported under the “theo.” column of the “Guitar-specific  $\beta$ ’s” heading in Table 4.1. This experiment was also repeated for more general inharmonicity contexts: we obtained log-inharmonicity estimates of all open-string notes belonging to the other guitars in the same class as the test guitar’s, as well as two-thirds of the test guitar’s open-string notes. As before, we performed three-fold cross validation on the test guitar’s recordings, and report error probabilities under the “theo.” column of the “Guitar-averaged  $\beta$ ’s” heading. This heading is named as such because the effect of learning from all three guitars is an “averaging” of the their open-string notes’ inharmonicities.

Our second experiment was an evaluation of our similarity metric using *empirical* inharmonicity trajectories, or inharmonicity regressions. For each of the guitars in our RWC subset, we obtained log-inharmonicity estimates of *all* labeled notes from two-thirds of its total recordings, then performed bisquare weighted regression against their pitches. String classification was performed on the remaining third using residual error minimization on the regressions, as discussed in Chapter 3. We repeat this using the same number of folds as the first experiment, and report error probabilities for each guitar under the “empi.” column of the “Guitar-specific  $\beta$ ’s” heading in 4.1. Again, we also evaluate regressions learned from all three guitars in the same class as the test guitar’s, using the same cross-validation configuration. These results are reported

under the "empi." column of the "Guitar-averaged  $\beta$ 's".

We quantified our system performance with error probabilities because this was the metric used by Barbancho in [6]. F1-score, which is the harmonic mean of the precision and recall, is also commonly used for classification performance measurement so we report it in Table 4.2 for reference.

Transcription Error Probabilities						
	Guitar-specific $\beta$ 's			Guitar-averaged $\beta$ 's		
Guitar	PCT	theo.	empi.	PCT	theo.	empi.
Classical CG091	0.04	0.02	0.04	0.09	0.02	0.04
Classical CG092	0.00	0.08	0.08	0.01	0.08	0.08
Classical CG093	0.01	0.07	0.08	0.00	0.08	0.09
Overall CG:	0.02	0.06	0.07	0.03	0.06	0.07
Acoustic AG111	0.00	0.01	0.04	0.06	0.00	0.04
Acoustic AG112	0.00	0.02	0.03	0.22	0.11	0.06
Acoustic AG113	0.00	0.03	0.06	0.04	0.04	0.08
Overall AG:	0.00	0.02	0.04	0.11	0.05	0.06
Electric EG131	0.00	0.20	0.22	0.20	0.19	0.21
Electric EG132	0.01	0.20	0.16	0.23	0.19	0.23
Electric EG133	0.00	0.12	0.07	0.32	0.16	0.24
Overall EG:	0.00	0.17	0.15	0.25	0.18	0.23

Table 4.1: Error probability comparisons between PCT, residual minimization with theoretical trajectories, and residual minimization with empirical trajectories. PCT results are taken from [6]; we've excluded error due to  $f_0$  estimation in their results.

We also display in Tables 4.3, 4.4, and 4.5 the string-wise F1-scores from this experiment. The left-hand column denotes the string number of the rows' F1-scores, and the right-hand column is the row-wise average of the three guitars in the class.

## 4.2 Tuning Compensation

For the next experiment, we learned inharmonicity regressions from the standard-tuned RWC electric guitar recordings, and predicted string classes on our alternate-tuned personal recordings. We evaluated string-wise F1-score performance both with and without the tuning compensation

Transcription F1-scores				
	Guitar-specific $\beta$ 's		Guitar-averaged $\beta$ 's	
Guitar	theo.	empi.	theo.	empi.
Classical CG091	0.98	0.96	0.98	0.96
Classical CG092	0.93	0.92	0.93	0.92
Classical CG093	0.92	0.92	0.91	0.91
Overall CG:	0.94	0.93	0.94	0.93
Acoustic AG111	0.99	0.97	1.00	0.96
Acoustic AG112	0.97	0.97	0.87	0.94
Acoustic AG113	0.97	0.94	0.96	0.92
Overall AG:	0.98	0.96	0.94	0.94
Electric EG131	0.78	0.78	0.79	0.78
Electric EG132	0.76	0.83	0.76	0.74
Electric EG133	0.89	0.93	0.79	0.76
Overall EG:	0.81	0.85	0.78	0.76

Table 4.2: F1-scores for theoretical and empirical log-inharmonicity trajectory classification. PCT column removed because F1-scores weren't reported in [6].

String No.	CG1	CG2	CG3	Overall CG
1	0.96	0.93	0.95	0.95
2	0.92	0.88	0.89	0.90
3	0.97	0.91	0.89	0.92
4	0.94	0.89	0.84	0.89
5	0.99	0.95	0.90	0.95
6	0.99	0.97	0.99	0.98

Table 4.3: String-wise F1-scores for the RWC classical guitars.

String No.	AG1	AG2	AG3	Overall AG
1	0.96	0.96	0.95	0.96
2	0.93	0.86	0.88	0.89
3	0.89	0.86	0.78	0.84
4	0.99	0.99	0.94	0.97
5	1.00	0.99	0.96	0.98
6	1.00	1.00	1.00	1.00

Table 4.4: String-wise F1-scores for the RWC acoustic guitars.

String No.	EG1	EG2	EG3	Overall EG
1	0.92	0.86	0.96	0.91
2	0.88	0.83	0.68	0.80
3	0.44	0.23	0.32	0.33
4	0.60	0.66	0.59	0.62
5	0.89	0.91	1.00	0.93
6	0.96	0.96	1.00	0.97

Table 4.5: String-wise F1-scores for the RWC electric guitars.

	Standard	"DADGAD"		"WSU"		"WSD"	
String No.	orig.	orig.	comp.	orig.	comp.	orig.	comp.
1	0.96	0.70	0.75	0.92	0.92	0.77	0.77
2	0.92	0.69	0.69	0.89	0.89	0.72	0.77
3	0.74	0.63	0.65	0.63	0.65	0.21	0.40
4	0.56	0.53	0.53	0.25	0.32	0.65	0.69
5	1.00	1.00	1.00	0.40	0.80	0.85	0.85
6	1.00	0.92	0.92	0.83	0.88	0.85	0.85
Overall:	0.86	0.74	0.76	0.65	0.73	0.67	0.71

Table 4.6: Best string-wise F1-scores for our Fender Telecaster at various tunings. Regressions trained on all three RWC electric guitars. WSU: whole-step up (from standard); WSD: whole-step down (from standard); orig.: no tuning compensation used in this trial; comp.: tuning compensation used in this trial. The "Overall" row displays average F1-score of each column.

factor in our system discussed in the previous chapter. Tuning compensation results are denoted by the "comp." column, and original results without compensation are denoted by the "orig." column. See Table 4.6. In standard tuning, strings 6 through 1 are tuned to E2, A2, D3, G3, B3, and E4. In so-called "dadgad" tuning, the open-string pitches are made to be D2, A2, D3, G3, A3, and D4 as these open pitches concatenated together form its name. Whole-step up and whole-step down are tunings in which each of the standard-tuned strings are respectively incremented or decremented by two semitones. For whole-step down, this results in D2, G2, C3, F3, A3, and D4. For whole-step up, we have F#2, B2, E3, A3, C#4, and F#4.



## Chapter 5: Discussion

In this chapter, we begin with an analysis of the previous chapter’s results, followed by consideration of factors that potentially threaten this work’s validity. We close with a discussion about future work.

### 5.1 Analysis

Encouragingly, our proposed inharmonicity similarity routine rivals the performance of PCT. Although we almost invariably performed worse than PCT did for test guitars to whose open-string log-inharmonicities we have access (the “Guitar-specific  $\beta$ ’s” heading in Tables 4.1 and 4.2), we achieve lower error probabilities for the most part when multiple guitars’ log-inharmonicities are considered (the “Guitar-averaged  $\beta$ ’s” heading). It seems this is a strength of our system; PCT shines with reliable guitar-specific inharmonicity information, but degrades faster and more variably when the reliability of that information decreases. Its worst case performance degradation occurs for the electric guitar, going from zero error to 25% error as the guitar-specific inharmonicity context is lost, but the error probability of trajectory residual classification, on the other hand, worsens only marginally at its worst case (also for the electric guitar), degrading by only 8 percentage points. PCT performance on guitar-averaged inharmonicities is also more erratic, with error probability variances for all three guitar classes about one order of magnitude greater than our residual classification. It appears that the residual classification approach trades peak performance capability for more stable results that are also more robust to different guitars’

inharmonicities.

Interestingly, classification using regressions on inharmonicity estimates collected from a single guitar generally performed worse than classification with the theoretical trajectories obtained from the open-string inharmonicity estimates of the same guitar. (Compare columns "theo." and "empi." under the "Guitar-specific  $\beta$ 's" heading). This was contrary to our hypothesis that the regressions would more accurately capture the empirical trajectory than would the theoretical one.

Insight into the error probabilities in Tables 4.1 is gained when considering the string-wise breakdowns of F1-scores in Tables 4.3, 4.4, and 4.5. We see that across all guitars, strings 2, 3, and 4 are consistently incorrectly transcribed. Strings 1, 5, and 6, contrastingly, are consistently accurate, so we know that any poor results in Table 4.1 stem from poor accuracy on these problematic "middle" strings that are physically surrounded by higher and lower strings on the instrument. We thus see a flaw in this residual classification approach: because of its inherently spatial classification decision (residual minimization), this system exhibits poor performance on strings whose trajectories share overlapping regions in inharmonicity-pitch space.

It's unclear why performance for electric guitars is invariably worse than for classicals and acoustics. To investigate, we visually compared the string-wise inharmonicity regressions of all guitars in each guitar class. As can be seen in Figures 5.1, 5.1, and 5.1, the classical and acoustic guitars' inharmonicity regressions remain fairly stable from guitar to guitar, whereas those of the electric guitars clearly exhibit more variability. EG regressions belonging to strings 3 and 4 overlap considerably, which is similar to what we see in the classicals (strings 1 and 4, 2 and 5, 3 and 6) and acoustics (strings 2 and 3) with some strings. But due to the large spread of the electrics' regressions, this overlap behavior causes frequent transcription errors. Indeed, the EG confusion matrices support this, and one which highlights this error trend is shown in Table 5.1. Example confusion matrices for the CGs and AGs are also shown. This overlapping error doesn't seem to affect the CGs and AGs nearly as much, since their regression variability is much smaller.



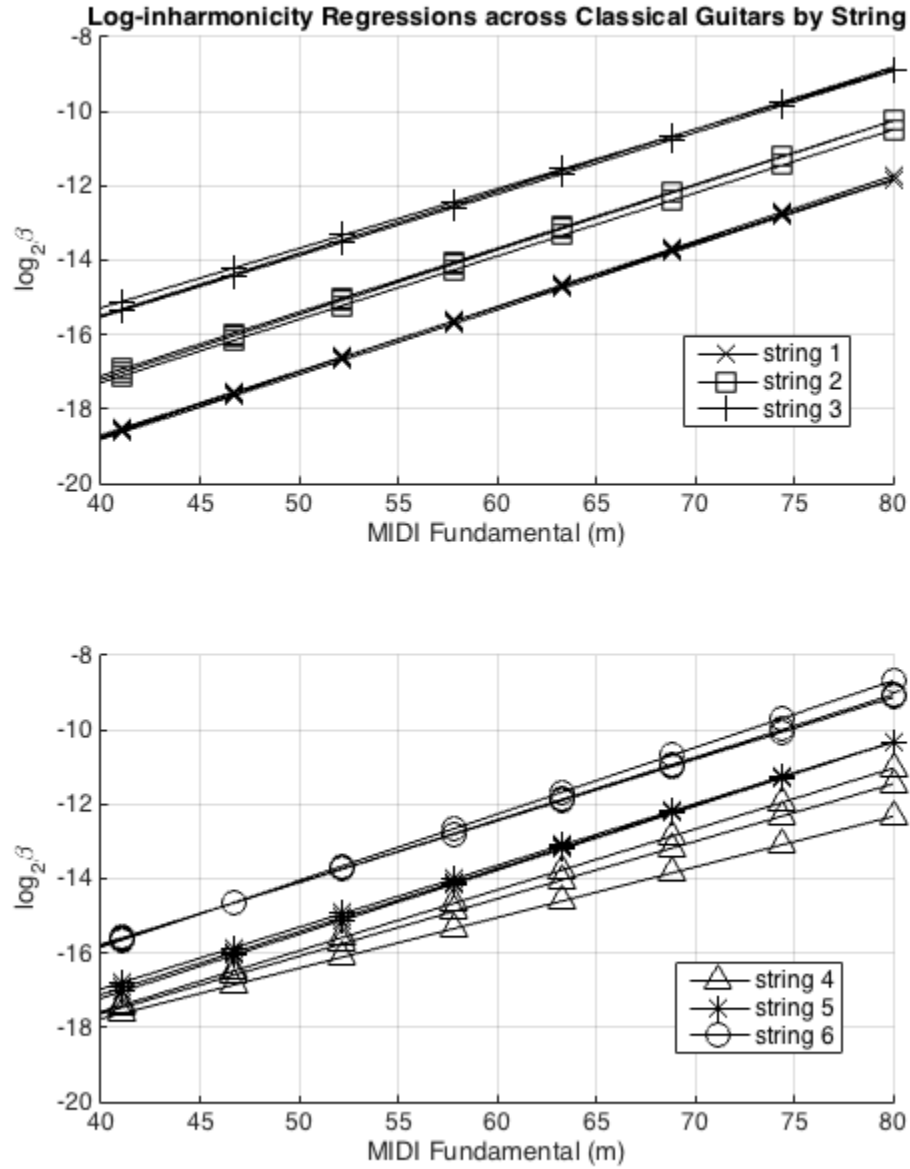


Figure 5.1: Superimposed string-wise regressions of all three RWC classical guitars. Top: strings 1, 2, and 3. Bottom: strings 4, 5, and 6. Separated for visibility. Axes scales identical for comparison. Observe the consistent placement and orientation despite varying guitar.

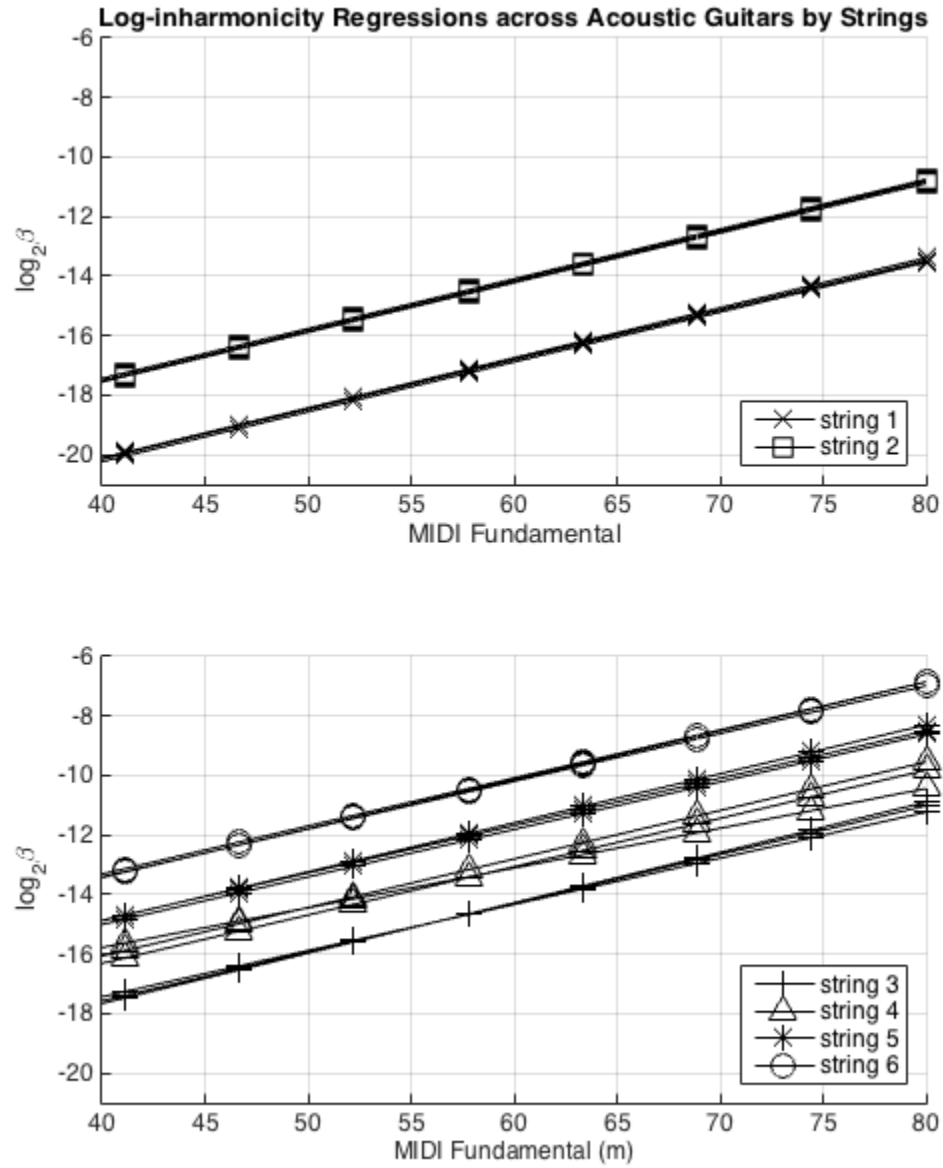


Figure 5.2:

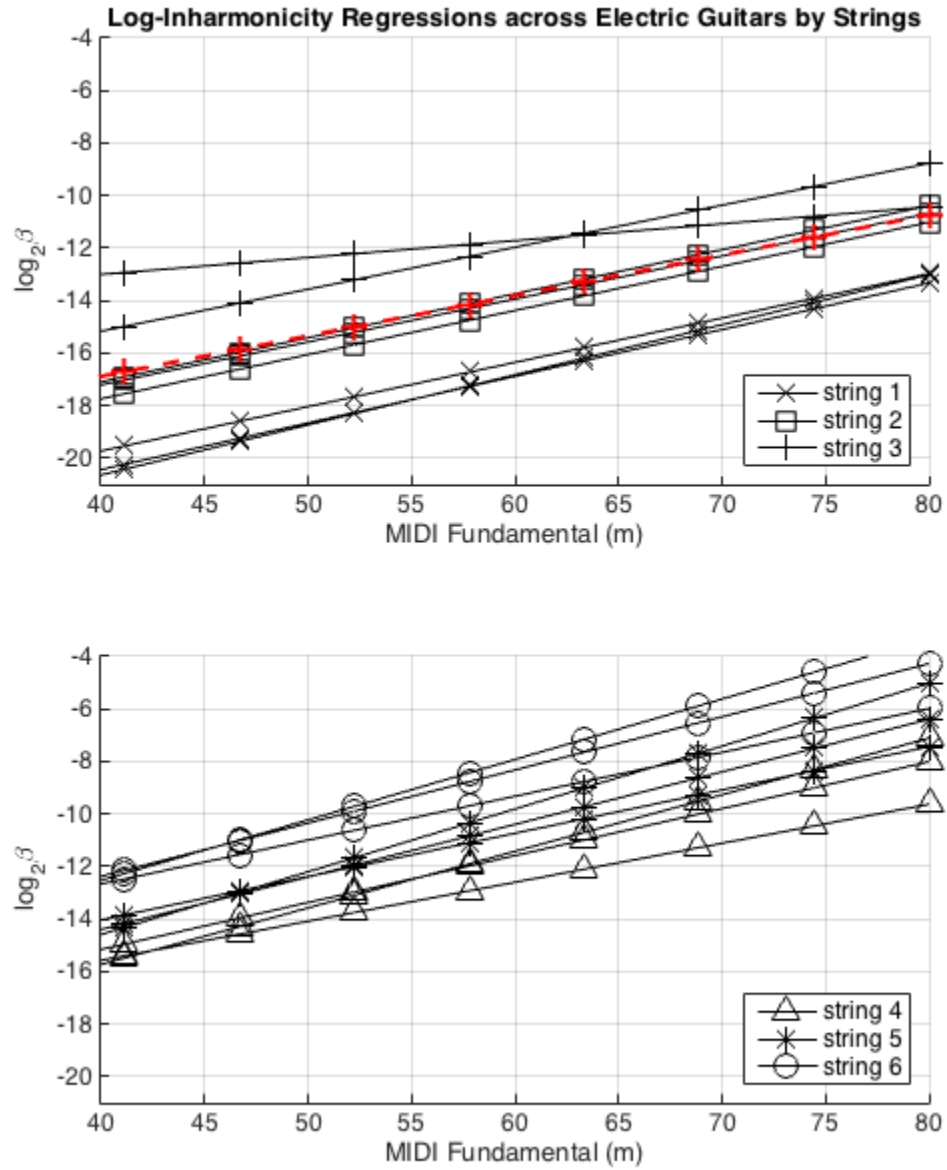


Figure 5.3:

	Classification					
Truth	1	2	3	4	5	6
1	<b>0.92</b>	0.08	0	0	0	0
2	0	<b>0.92</b>	0.08	0	0	0
3	0	0.08	<b>0.38</b>	0.54	0	0
4	0	0	0.08	<b>0.62</b>	0.30	0
5	0	0	0	0	<b>0.77</b>	0.23
6	0	0	0	0	0	<b>1.00</b>

Table 5.1: Confusion matrix for one classification trial with EG131. Strings 3 and 4, as they are here, were frequently misclassified because of their regressions’ poor discriminability.

Further analysis revealed that inconsistent inharmonicity estimates were responsible for the electric guitars’ regression variability. We examined the quality of fit of string regressions obtained from all recordings of each of the 3 guitars from each of the 3 types, and found that the classics’ and acoustics’ range of determination coefficients was substantially smaller than the electrics’. Regressions for CG091, AG111, and EG131 are shown in Figure 5.1 as an example of this; CG091’s and AG111’s best and worst strings’  $r^2$  coefficients differ by only about one-hundredth, but the  $r^2$  difference for EG131 is an enormous 0.87.

Speculating why this might be the case, we wondered if the transduction of the EG pickups somehow played a role in obscuring reliable inharmonicity estimation – either through their transfer function, possible noise introduction, or some non-linear characteristic. If this were the case, it would follow that higher-intensity performances would engage with the obscuring pickup phenomenon to a greater extent and produce measurably more variable inharmonicity estimates than lower-intensity performances. Since one of the RWC musicality parameters that varies for each featured guitar is dynamic level of the performance, we were able to separate ”forte” and ”piano” performances of the same guitar and analyze this interaction between degree of pickup transduction and inharmonicity estimation variability. For each guitar, and for each string, we collected note performances by common dynamic levels (piano, mezzo, and forte) and regressed their log-inharmonicity against their fundamentals, as usual. We then plotted statistics of the log-inharmonicity’s residuals against their string’s regression. Figure 5.1 illustrates these

### Best-Worst Fit Comparisons for Log-Inharmonicity Regressions

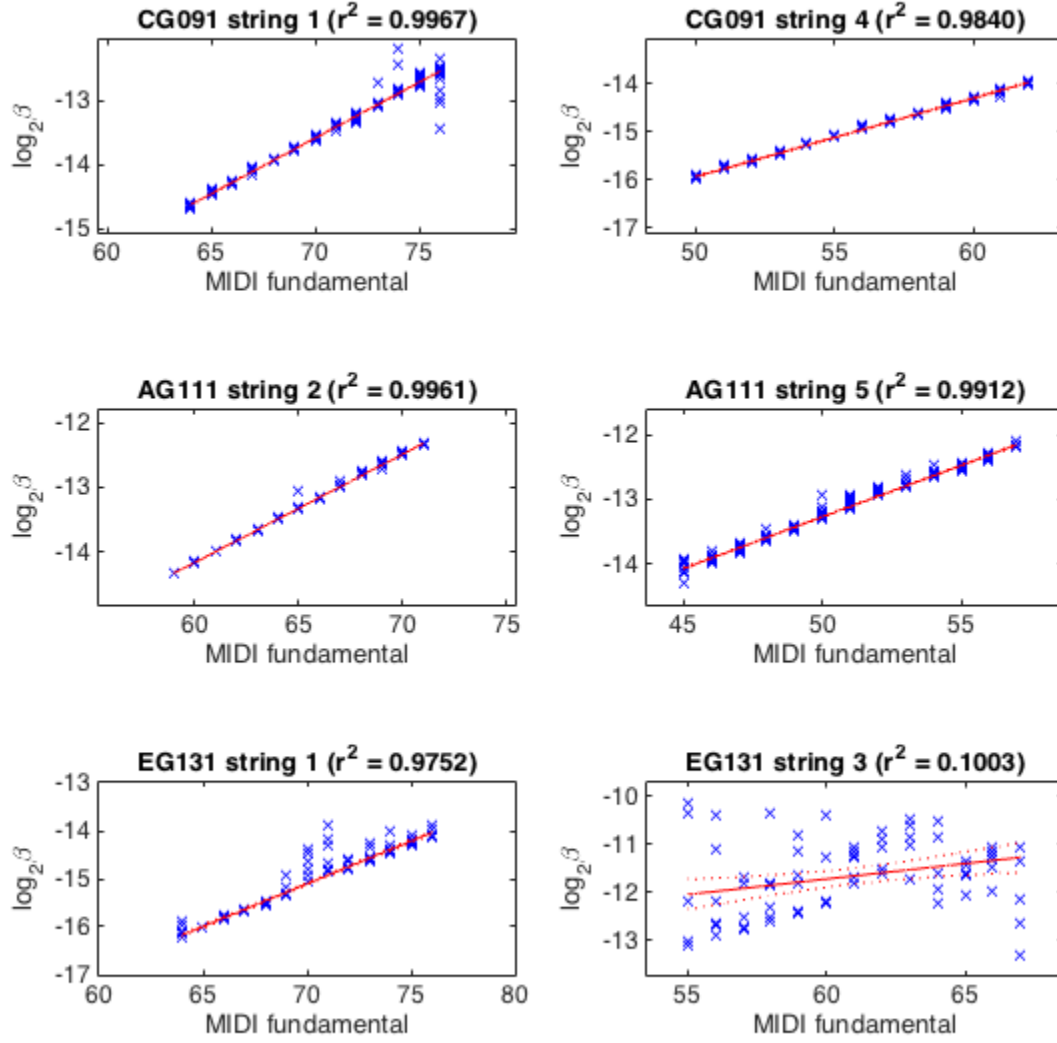


Figure 5.4:

statistics for EG131; other electric guitars are omitted because their statistics were similar. We

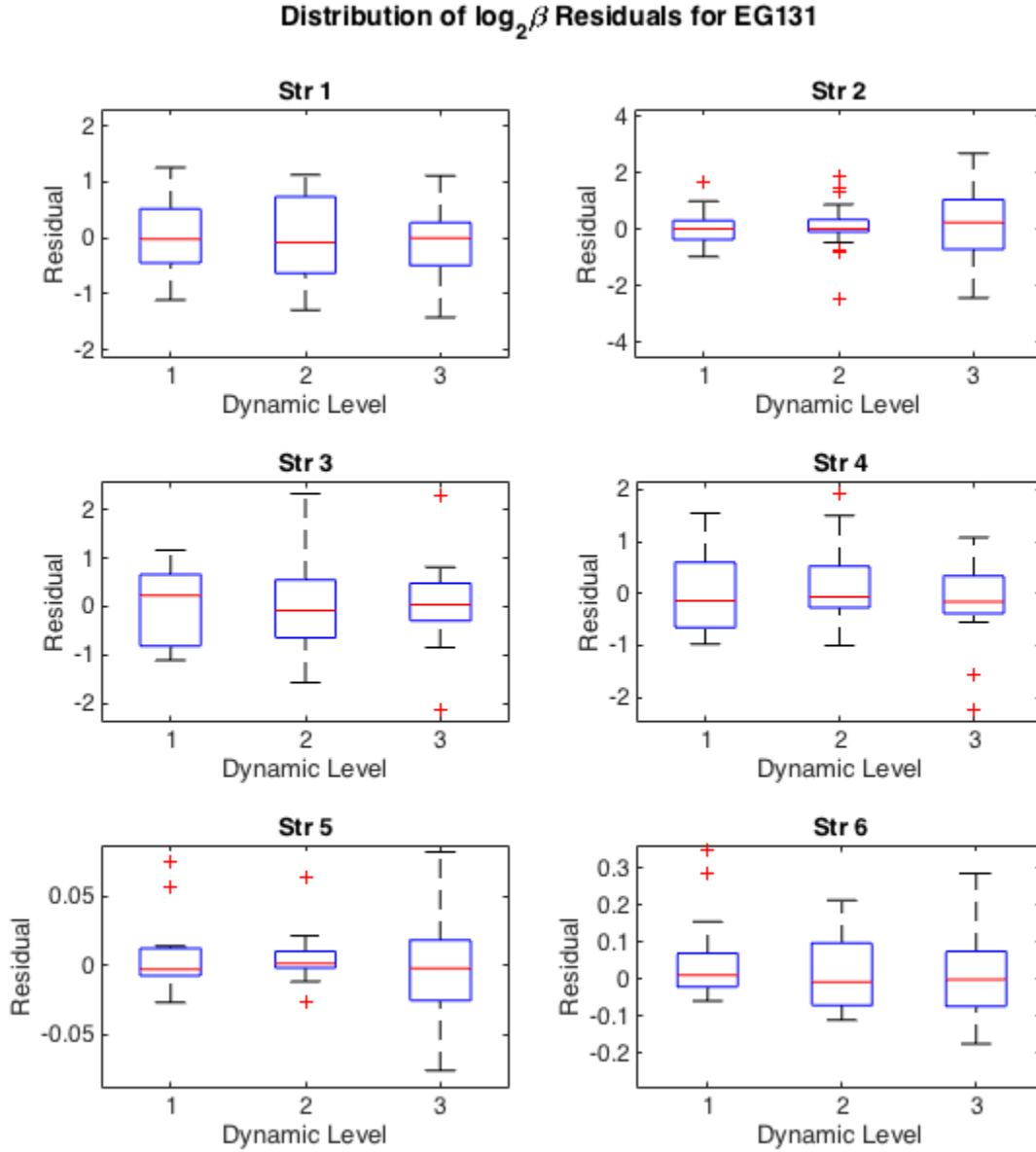


Figure 5.5: Boxplots of log-inharmonicity regression residuals for EG131 by dynamic level. Group 1: piano, Group 2: mezzo, Group 3: forte. Red center horizontal lines are medians; top and bottom edges of blue boxes represent upper and lower quartiles; whiskers depict full range of data excluding outliers; red plus signs denote outliers. Vertical axes not constant throughout plots; we're highlighting relative distributions of the boxplots here.

expected that variability would increase with dynamics, but unfortunately the data didn't support

this hypothesis. Interquartile ranges didn't exhibit any clear trend with respect to performance intensity, nor was the magnitude of their variation noteworthy. The cause of the inharmonicity unreliability is therefore probably linked to some spectral characteristic of the electric guitars, e.g. considerably more non-tonal peaks which the MAT algorithm erroneously selects as partials, or some aspect of the pickup's frequency response that's throwing off the MAT algorithm, etc.

## 5.2 Threats to Validity

It's important to highlight factors that potentially undermine the claims of this work. Our system currently operates only on clean (i.e. no effects processing), isolated, monophonic guitar recordings, which are obviously not representative of the majority of guitar audio. Multiple layers of effects processing are commonly used by performers of the instrument to achieve distinctive, creative sounds. Guitarists often play in concert with other musicians, and frequently aren't the focal point of the song, buried in their companion's soundscapes. Moreover, guitar passages are rarely in their entirety strictly monophonic. It should be noted that these are large obstacles that separate our RWC subset from actual guitar performances, and thus the implications of our transcription results should be considerably tempered. Noteworthy though, is the extent to which other teams have overcome some of these factors. Barbancho, Abesser, Kehling, and Dittmar have implemented polyphonic capabilities into their systems with good success, so polyphony is perhaps a smaller barrier than the others. Maezawa's system achieved modest performance on RWC recordings that featured guitar passages combined with backing tracks.

Additionally, we assumed perfect fundamental pitch estimation in this work. Our decision to simply use the RWC labels stemmed from convenience and scope; the focus of this work was simply evaluation of a specific approach to examining notes' inharmonicity for tablature transcription, and having not to worry about the reliability of a pitch detection routine certainly expedited this thesis. That being said, it should be acknowledged that the reliability of one's pitch estimator is a sure factor in the translation of this work's results to the real world.

Furthermore, the MAT inharmonicity estimation routine we used was different from Barbancho's. The credibility of the comparison between our results and theirs is thus somewhat compromised. It's quite possible that any performance gains observed in our implementation are due in part to potentially superior inharmonicity extraction. This is, unfortunately, a factor we didn't control for, and if time had allowed we should have replicated Barbancho's system and substituted their inharmonicity estimation step with MAT to produce results that were as comparable as possible.

## 5.3 Future Work

A potential area for run-time performance enhancement for tablature transcription systems is the use of musical context. Musical passages on the guitar tend to exhibit continuity; it's usually the case that a riff is played in a localized area of the fretboard, for example, instead of jumping around the length of the guitar neck. It's empirically more natural for a performer to be motion conservative in this sense, as it's easier to play more notes quicker and it the piece sounds more uniform. Transcription systems could apply these performance-borne simplifications to their output to potentially improve accuracy, say, when a stray note is transcribed a plausible twelve frets away from the rest of those in the riff.

Necessary to do this, though, is a measure of objective confidence in the transcription. Without this, there'd be no way to determine whether it was the stray note or the majority of the riff which was correctly transcribed. We investigated the amenability of our log-inharmonicity residual-minimizing classification scheme to reliable confidence measurements, and reassuringly found encouraging results. Figure 5.3 displays the probability, for each of the nine RWC guitars, of our system producing a correct fretboard transcription given increasingly wide windows in which we consider a note's inharmonicity estimate to have fallen. The inverse trend between residual window width and correctness probability shows that we can reliably infer some degree of confidence in a note's transcription. This can be leveraged in a post-processing stage to further



refine tablature with isolated erroneous, yet plausible, transcriptions.

Another dimension we didn't consider was varying the regression polynomial. It would be interesting to see how differently, if at all, the system would perform if the log inharmonicity transformation was forgone and an exponential fit was explored instead. Perhaps the ratio between inharmonicity estimation noise and the exponentials' residuals would be appreciably different to affect some change in the transcription results. This should be investigated in future inharmonicity-based transcription research.

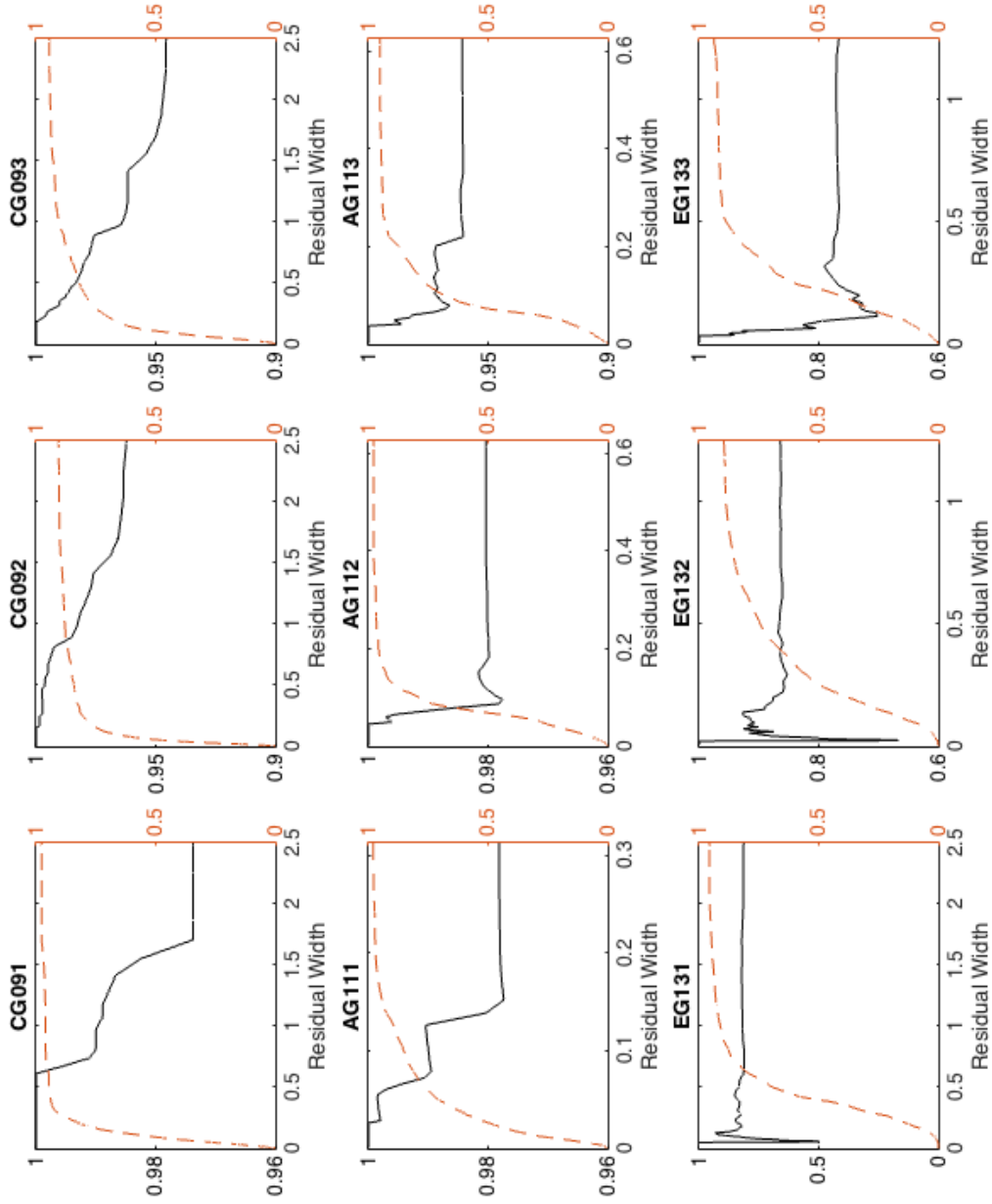


Figure 5.6: Solid black lines: probability of correctly transcribing a note given the residual width under which the note's log-inharmonicity falls, measured by the left-hand-side vertical axes. Dashed orange lines: CDF of the log-inharmonicity residuals, measured by the right-hand-side vertical axis.

## Chapter 6: Conclusion

We began by introducing the guitar and an essential component of its musicality: the overlapping pitch ranges of its strings. We related this to tablature and its ability to uniquely specify fretboard positions for musical passages, then highlighted its popularity and its tedious manual annotation process. This led us to introduce automatic transcription systems and their potential benefits for music students, and we reviewed previous work on this rich topic.

We paid special attention to Barbancho [6], who exploited only one feature known as inharmonicity in development of a successful tab transcription system using only audio. We explained how inharmonicity affects the degree of upward skew of a note’s partials, then discussed the intuition behind its discriminative power before introducing our novel approach: characterizing strings as log-inharmonicity trajectories and regressions.

After reiterating one of Barbancho’s key realizations – that inharmonicity along a given string was deterministic – we showed that the logarithm of these trajectories was linear with respect to MIDI pitch. We consequently proposed characterizing guitar strings by their log-inharmonicity lines and performing classification (and therefore implicitly, tablature transcription) by assigning to unknown notes the strings whose lines best approximated the notes’ log-inharmonicity. Lastly, we derived the effect of a tuning change on these lines, and arrived at a general tuning-compensation feature applicable to any inharmonicity-based system.

Next, we introduced our experimental datasets – a subset of the RWC music instruments database and a couple personal guitar recordings – and outlined the experiments we conducted. We first performed benchmark comparisons against Barbancho’s similar inharmonicity-based

method, and then we assessed our tuning compensation feature’s transcription performance for alternately-tuned guitars. Both evaluation areas were promising; our RWC tab transcription performance rivaled that of Barbancho’s at 0.92, 0.94, and 0.79 for the classical, acoustic, and electric guitars respectively, while our average system accuracy on alternate-tuned guitars improved by 7% compared to performance with no tuning compensation.

We wrapped up with a more detailed discussion of our experiment results, and an analysis of some of its less encouraging portions. A cursory investigation of the cause for the poorer electric guitar transcription performance revealed that it couldn’t be the electric guitars’ pickups that were responsible for impeding reliable inharmonicity estimation; it must rather be spectral nuances that confuse the inharmonicity estimation algorithm. We closed with limitations and future scope. More work should be done for a thorough appraisal of our system’s merit, but results obtained thus far support the promise of our novel automatic tablature transcription approach.

## Bibliography

- [1]
- [2] .
- [3] .
- [4] J. Abeßer. Automatic string detection for bass guitar and electric guitar. In *From Sounds to Music and Emotions - 9th International Symposium, CMMR 2012*, pages 333–352, London, UK, June 2012.
- [5] I. Barbancho, A. Barbancho, L. Tardón, and S. Sammartino. Pitch and played string estimation in classic and acoustic guitars. In *Proceedings of the 126th Audio Engineering Society (AES) Convention*, Munich, Germany, 2009.
- [6] I. Barbancho, L. Tardón, S. Sammartino, and A. Barbancho. Inharmonicity-based method for the automatic generation of guitar tablature. In *Proceedings of the IEEE Transactions on Audio, Speech, and Language Processing*, pages 1857–1868, August 2012.
- [7] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler. A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, 13(5):1035–1047, 2005.
- [8] G. Burlet and I. Fujinaga. Robotaba guitar tablature transcription framework. In *Proceedings of the IEEE Transactions on Speech and Language Processing*, Curitiba, Brazil, November 2013.
- [9] C. Dittmar, A. Mäanchen, and J. Abeßer. Real-time guitar string detection for music edu-

- cation software. In *Proceedings of the 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pages 1–4, 2013.
- [10] S. Dixon. Onset detection revisited. In *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06)*, Montreal, Canada, September 2006.
- [11] H. Fletcher, E. Blackham, and R. Stratton. Quality of piano tones. *Journal of the Acoustical Society of America*, 34(6):749–761, 1962.
- [12] A. Galembo. Method for objective evaluation of the sound quality in piano trebles (in russian). In *Proc. of 18th Acoust. Conf. of Czechoslovakia*, 1979.
- [13] A. Galembo. *The quality of piano tones (in Russian)*. Legkaya Industria Press, Moscow, 1987.
- [14] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Music genre database and musical instrument sound database. In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pages 229–230, October 2003.
- [15] M. Hodgkinson, J. Wang, J. Timoney, and V. Lazzarini. Handling inharmonic series with median-adjustive trajectories. In *Proc. of the 12th Int. Conference on Digital Audio Effects (DAFx-09)*, Como, Italy, September 2009.
- [16] C. Kehling, J. Abeßer, C. Dittmar, and G. Schuller. Automatic tablature transcription of electric guitar recordings by estimation of score- and instrument-related parameters. In *Proc. of the 17th Int. Conference on Digital Audio Effects (DAFx-14)*, Erlangen, Germany, September 2014.
- [17] R. Macrae and S. Dixon. A guitar tablature score follower. In *Proc. Int. Conf. Multimedia Expo*, pages 725–726, 2010.
- [18] A. Radisavljevic and P. Driessen. Path difference learning for guitar fingering problem. In *Proc. Int. Comput. Music Conf. (ICMC04)*, 2004.
- [19] J. Rauhala, H. M. Lehtonen, and V. Välimäki. Fast automatic inharmonicity estimation

June 16, 2017

DRAFT

algorithm. *Journal of the Acoustical Society of America*, 121(5):184–189, 2007.